Performance Evaluation of Scalable Multi-cell On-Demand Broadcast Protocols

A Thesis Submitted to the
College of Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the degree of Master of Science
in the Department of Computer Science
University of Saskatchewan
Saskatoon

By
Yuntao Mei

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science
176 Thorvaldson Building
110 Science Place
University of Saskatchewan
Saskatoon, Saskatchewan
Canada
S7N 5C9

Abstract

As mobile data service becomes popular in today's mobile network, the data traffic burden irrevocably increases. LTE 4G, as the next-generation mobile technology, provides high data rates and improved spectral efficiency for data transmission. Currently in the mobile network, mobile data service solely relies on the point-to-point unicast transmission. In the ever-evolving 4G mobile network, mobile broadcast may serve as a supplemental means of pushing mobile data content from the data server to the mobile user devices. As part of the LTE 4G specifications, the mobile broadcast technology referred to as eMBMS is designed for supporting the mobile data service. From eMBMS, SFN broadcast transmission scheme allows data broadcasting to be synchronized in all cells of a defined core network area. LTE 4G also enables single-cell broadcast scheme in which data broadcasting is taking place independently in every cell.

In this thesis, besides SFN or single-cell broadcast transmission, a hybrid broadcast transmission scheme in which SFN and single-cell broadcast transmission are used interchangeably in the same network based on the network conditions is proposed. For on-demand data service, the pull-based scheduling protocols from previous work are originally designed to work in a single-cell case scenario. With slight modifications, the batching/cbd protocol can be adapted for multi-cell data service. A new combined scheduling protocol, that is cyclic/cd,fft protocol, is devised as the second candidate for multi-cell data transmission scheduling. Based on the three broadcast transmission schemes and the two broadcast scheduling protocols, six mobile broadcast protocols are proposed. The mobile broadcast models, which correspond to the six mobile broadcast protocols, are evaluated by analysis and simulation experiment. By analysis, the cost equations are derived for calculating average server bandwidth, average client delay and maximum client delay of the mobile broadcast models. In the experiment, the input parameters of broadcast test models are assessed one at a time. The experimental results show that the hybrid broadcast transmission together with cyclic/cd,fft protocol would provide the best server bandwidth performance and the SFN broadcast transmission together with batching/cbd protocol provides the best average delay performance.

Acknowledgements

I would like to thank my supervisor, Dr. Derek Eager, for his guidance and support throughout my graduate studies. I greatly appreciate the time and effort he spent on helping me find the research topic, discussing my research work and correcting my thesis. My current academic achievements cannot be accomplished without his instruction. It has been a great experience working under Dr. Derek Eager's supervision.

I would also like to thank my thesis committee members: Dr. Dwight Makaroff, Dr. Mark Keil and Dr. Raj Srinivasan, for their helpful comments and suggestions.

Finally, I want to offer my special thanks to my parents, whose constant love and care have helped me so much.

Contents

List of Tables

List of Figures

List of Abbreviations

1G	First Generation (mobile communication system)
2G	Second Generation (mobile communication system)
3G	Third Generation (mobile communication system)
3GPP	Third Generation Partnership Project
4G	Fourth Generation (mobile communication system)
batching/cbd	batching with constant batching delay
CDN	Content Delivery Network
CoMP	Coordinated Multipoint
CSI	Channel Status Information
DVB-H	Digital Video Broadcasting - Handheld
cyclic/cd,bot	cyclic with constant delay, bounded on time
cyclic/cd,fft	cyclic with constant delay, full-file transmission
eMBMS	Evolved Multimedia Broadcast/Multicast Service
FCFS	First Come First Serve
HSPA	High Speed Packet Access
IMT-Advanced	International Mobile Telecommunications-Advanced
LMF	Longest Wait First
LTE	Long Term Evolution
MBMS	Multimedia Broadcast/Multicast Service
MBSFN	Multicast-broadcast Single Frequency Network
MCE	Multiple-cell Coordination Entity
MNO	Mobile Network Operator
MRF	Most Requested First
PLMF	Preemptive Longest Wait First
RMSD	Root-mean-square Deviation
RNC	Radio Network Controller
SFN	Single Frequency Network
SINR	Signal to Interference Noise Ratio
SRST	Shortest Remaining Service Time
SSTF	Shortest Service Time First
UMTS	Universal Mobile Telecommunications System

Chapter 1

Introduction

In the past decade, 3G broadband networks were widely deployed around the world and the mobile network has begun to offer a wide range of mobile data service besides the traditional voice communication service. In the recent few years, sales of smartphones, mobile PCs and tablets has boomed in the mobile market in many countries. The popularity of the large-screen mobile devices has driven the substantial growth of the broadband data service subscriptions from mobile service users. In today's mobile network, the volume of mobile data consumption continues to rise and the ever-growing mobile data traffic has imposed a strain on the mobile data networks. For the MNOs (Mobile Network Operators), it has become necessary to bring in innovative solutions in response to the trend of increasing demand for mobile data. From the perspective of the current mobile industry, one important goal is to develop new mobile technology that further expands the capacity of mobile data transfers from limited radio bandwidth resources.

Following the evolution path of the 3G technologies, LTE (Long Term Evolution) and its evolution LTE-Advanced (LTE-A) are generally considered as the next-generation cellular technology [38]. The LTE project was initiated by Third Generation Partnership Project (3GPP) as a collaborative effort to achieve 4G wireless data communication standard. LTE technology has been developed with the major design focus on increasing the capacity and speed for mobile broadband data transmission in both uplink and downlink. LTE also maintains backwards compatibility with the current mobile telephony technology like GSM and HSPA. This allows MNOs to adopt LTE on the existing network infrastructure without too much cost [1]. Since the first release (LTE release 8) in March 2008, LTE has been gradually updated and at each time introduced new enhanced features for improving data transmission performance. LTE release 10, also known as LTE-Advanced, offers high data rate, improved spectral efficiency, and reduced latency. LTE-Advanced is the first LTE release that meets the requirements of IMT-Advanced standard and it is regarded as LTE 4G [21, 38]. The successive LTE release, release 11, had redesigned the core network architecture and air interface which further increased the spectral efficiency and expanded the data rate capacity [8]. An increase in the spectral efficiency means given the same quality of data service, the data server becomes capable of serving more clients, or for the same number of requesting clients, the throughput for each client increases. Compared with 3G technology, the current LTE 4G technology is able to provide higher quality broadband data service with minimized bandwidth resources in an efficient manner.

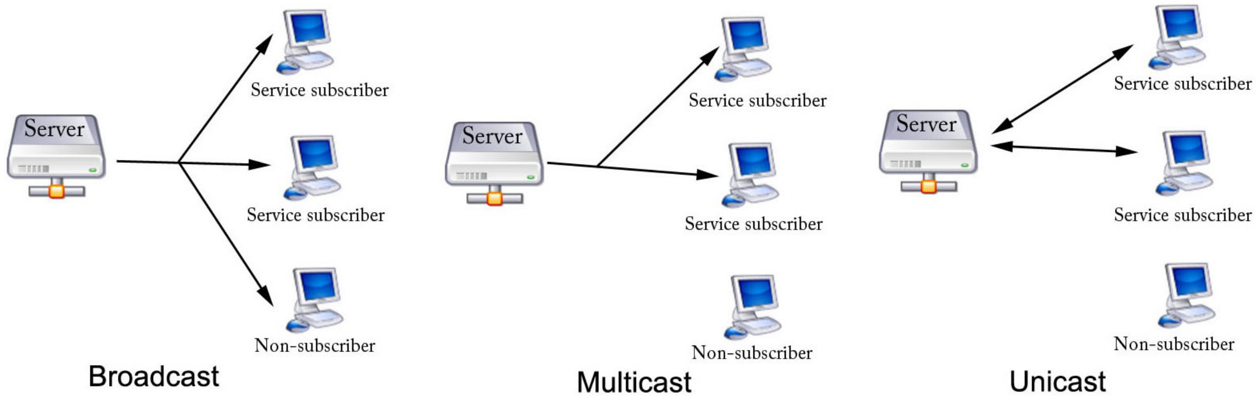


Figure 1.1: Broadcast, multicast and unicast

There are three ways of pushing data content from a server to the end-user devices: broadcast, multicast and unicast [29]. Terrestrial radio and television are the typical examples of the broadcast networks. For broadcast, the media data is transferred from the data server to the end-user devices on a single unidirectional channel shared by all listeners. All end-user devices within the coverage of the terrestrial radio or television networks receive the broadcast service. Multicast systems allow the server to deliver the data content only to those end-user devices that have joined the service. Since only the designated listeners are expected to receive the data service, the multicast server should not only store and transmit the data content, but also keep a record of the certain group of listeners that are qualified for the multicast service. For unicast communication, the system provides a bidirectional link between the data server and each end-user device. The seamless connection in unicast would allow real-time voice and video communication to take place.

Prior to LTE technology, unicast transmission had been the primary means of delivering data in the mobile networks. As opposed to unicast, broadcast/multicast is an efficient solution for distributing the same data content to a large number of recipients and has been used in many data transmission applications. However, in the 3G mobile network, the broadcast transmission has not been commercially utilized, partly due to the fact that the broadcast service enabled by 3G was only limited to fixed schedule and the benefit of using 3G broadcast transmission might not be able to redeem the cost of updating the unicast-based network infrastructure [14, 23]. Nowadays, the mobile data service demands more and more bandwidth resources for the growing data requests. In order to alleviate the data traffic burden, the broadcast approach has started to draw attention from the mobile industry as one of the viable solutions [33]. As the milestone development of mobile communication technology, the LTE-Advanced incorporated the mobile broadband broadcast transmission, which is known as eMBMS (evolved Multimedia Broadcast/Multicast Service). From the LTE-Advanced, eMBMS has become a part of the LTE 4G standard specifications and it has been maintained and refined in later LTE releases [29].

As the broadcast technology for LTE 4G, eMBMS has two main advantages over other competing technologies such as DVB-H: performance and cost. The broadcast service provided by eMBMS is intrinsically based on the LTE 4G infrastructure. Thus, it makes effective use of all the performance enhancements that LTE 4G network provides. The current major enhancements include high bit rates, flexible spectral usage and the deployment of SFN (Single Frequency Network). Performance enhancements from LTE 4G enable eMBMS to achieve improved performance for broadcast service [22, 29]. SFN broadcast transmission is particularly useful for large-scale data dissemination. It allows the same radio signals to be synchronized and simultaneously transmitted over the common frequency band to the end users within a defined mobile network area. During LTE data transmission, the base stations may use different frequency bands for the uplink and downlink traffic, or use the same frequencies for both uplink and downlink, alternating in time between the uplink and downlink traffic [24]. The use of the addressable time-frequency blocks, which consist of multiple consecutive sub-carriers for the duration of one time slot, would also facilitates synchronization of the mobile data transmissions among user devices [24]. For any 3G mobile network, the LTE broadcast service would not incur any additional hardware expenditure other than the deployment cost of the LTE 4G network infrastructure [22]. If a 4G network is in use, then LTE broadcast is expected to have lower operational cost than the current alternative mobile broadcast technologies which normally require additional hardware upgrades.

After LTE broadcast was fully integrated into 4G technology, some white papers[1,2] predicted that LTE broadcast would mark a profound shift in the mobile data service paradigm from the point-to-point transmission to the point-to-multipoint transmission. The white papers argued that because LTE 4G transmission/reception devices have been installed with compatible chipsets and middleware for broadcast, LTE RAN (Radio Access Network) would not require hardware changes for broadcast transmission and LTE broadcast can be made easily accessible to both MNOs and data service subscribers. Once LTE broadcast becomes active in mobile networks, MNOs can create more revenue opportunities by implementing a variety of broadcasting applications for data service subscribers. Since LTE broadcast makes more efficient use of valuable bandwidth resources, the mobile data service subscribers can be provided with higher quality data service with enhanced user experience. The white papers also proposed some use cases where LTE broadcast can be deployed for delivering the same data content to a large number of recipients. In these use cases, LTE broadcast is expected to offer more efficient distribution of media data than the point-to-point unicast transmission.

The LTE broadcast use cases are categorized into three types of data service.[1,2] The first is *live streaming service*, in which mobile data recipients listen on an LTE broadcast channel in order to receive the scheduled broadcast of audio or video content. The second is *on-demand broadcast streaming*. Instead of using a fixed schedule, the broadcast decision is based on a consensus of on-demand requests. Once the broadcast decision

[1] http://www.expway.com/wp-content/uploads/White-Paper-14-LTE-Broadcast-Use-Cases-final.pdf, access 14-July-2016
[2] https://www.qualcomm.com/documents/content-all-potential-lte-broadcastembms-white-paper, access 14-July-2016

is made, the popular data content like audio or video streams is transferred in their original order through the LTE broadcast channel to the mobile devices with low playback delay. For every user request, the mobile device establishes a connection and exchanges control information with the data server on a separate unicast channel. The typical example for *on-demand broadcast streaming* would be broadcasting YouTube or Netflix videos to mobile devices based on the requests. The third is *on-demand broadcast download*. Unlike *on-demand streaming*, after the broadcast decision is made based on the on-demand requests, the data content is transferred through the LTE broadcast channel to mobile devices with tolerable delay. The data content received by the user might not be in its original order. The typical examples for *on-demand broadcast download* may include mobile preloading of videos or other content that the user may wish to view later, off-peak media delivery, mobile software/app/firmware updates.

There are different designs for how LTE broadcast can be efficiently utilized in 4G networks. One design is that an LTE mobile system can solely rely on either broadcast transmission over an MBSFN (Multicast-broadcast Broadcast Single Frequency Network) area or broadcast transmission in the single cell [22]. For SFN broadcast transmission, the whole multi-cell region is treated as a single cell and the broadcast of the same data content is synchronized across all of the region's cells. For single-cell broadcast transmission, the broadcast of the same data content only takes place independently in every individual cell. This mobile broadcast design is most suitable for a mobile network area where there are heavy request demands for the same data content and the data requests are highly predictable. For example, in the places like a large sports event venue or an airport, mobile data traffic always tends to be heavy. Certain data content, such as the live commentary of the sports event or the flight schedule information, is expected to be frequently requested. The data server receives the requests on demand and broadcasts the requested data content to the designated mobile cells by means of single-cell broadcast transmission or to a multi-cell network region by means of SFN broadcast transmission.

Another design for mobile broadcast is that the LTE broadcast approach serves as a supplemental means of pushing mobile data content from the data server to the mobile user devices in complement to the point-to-point unicast communication [34]. In accord with this design, unicast transmission, SFN broadcast transmission and single-cell broadcast transmission are all enabled in the mobile system and one of these schemes is adaptively selected to use for the mobile data service based on the network conditions. The pertinent networks conditions may include the number of outstanding data requests for the same data content in the cell and the percentage of cells in the broadcast network that have at least one outstanding data request for the same data content. With the help of feedback mechanisms in the LTE 4G network, end user information can be obtained through a polling technique [22].

To carry out polling, a feedback channel is allocated between the user's mobile device and its base station. In every individual cell, the base station can keep track of the number of outstanding data requests for the same data content and forward it to the data server. In the mobile network, the information on the number of cells with at least one outstanding data request for the same data content can be gathered at the data server from the feedback of the base stations. Based on the feedback information, the data server can calculate the radio bandwidth required for unicast transmission, single-cell broadcast transmission, and SFN broadcast transmission. By comparing the projected results from these calculations, the data server is able to select the broadcast transmission scheme that provides the optimal use of the available bandwidth resources for the same data service [22].

Since the network conditions would change over time, the data server needs to periodically collect feedback information and update the broadcast transmission decision accordingly in every cell. If the number of outstanding data requests within a cell for the same data content is detected to be below a threshold, then unicast transmission should be selected for mobile data service in the cell. Otherwise, broadcast should be used for the data service. SFN broadcast transmission should be applied in all cells in place of the other alternative transmission schemes only when the number of cells with at least one outstanding data request for the same data content exceeds a threshold. With this adaptive transmission design, the maximum amount of bandwidth resources required for data transmission should only be determined by the guaranteed quality of data service, rather than the scalable number of outstanding requests in the same broadcast channel.

1.1 Thesis Motivation and Approach

With the advent of eMBMS, mobile broadcast may become an applicable approach for mobile data service. The interest of this research is placed on the application of scalable multi-cell on-demand broadcast, which is enabled by eMBMS from LTE 4G and corresponds to many use cases of mobile data service. Various pull-based broadcast protocols have been studied in the previous works [15, 48], but not all of them can effectively be adapted to work in the multi-cell mobile environment. The mobile broadcast protocol should not only work properly in the multi-cell mobile environment, but also provide performance benefits, such as reduced bandwidth requirement and minimized service delay time, in the data service.

In this research, a mobile broadcast protocol is considered to be composed of two parts, a suitable mobile broadcast transmission scheme and an efficient broadcast scheduling protocol. The mobile broadcast transmission schemes supported in LTE 4G are the SFN broadcast transmission and single-cell broadcast transmission. Based on these two basic broadcast transmission schemes, a new hybrid broadcast transmission, which heuristically combines both SFN and single-cell broadcast transmission, is further proposed as the solution in dynamic network conditions. For the efficient broadcast scheduling protocol, a batching protocol and a cyclic combined protocol are proposed as candidates, which are capable of efficiently responding to different request arrival patterns. Both protocols are designed specifically to work in the multi-cell mobile environment. To construct a mobile broadcast protocol, one of the three broadcast transmission schemes can be combined with one of the two broadcast scheduling protocols. The different possible combinations result in six mobile broadcast protocols whose performance can be assessed and compared.

1.2 Thesis Objectives

Thesis objectives are as follows:

- to explore mobile broadcast transmission schemes that are enabled in LTE 4G mobile network,
- to propose the design of the broadcast scheduling protocols that are suited for data service in the multi-cell mobile environment,
- to propose mobile broadcast protocols
- to develop analytic performance models for the mobile broadcast protocols,
- to assess the performance of the mobile broadcast protocols, and the accuracy of the analytic models, through simulation experiment.

1.3 Thesis Findings

Based on three broadcast transmission schemes and two broadcast scheduling protocols, six different mobile broadcast protocols are proposed and they are designed to work in various mobile network conditions. The three broadcast transmission schemes include two basic transmission schemes inherently supported by LTE 4G, which are single-cell broadcast transmission and SFN broadcast transmission, and a new hybrid broadcast transmission. The two broadcast scheduling protocols are the batching/cbd protocol proposed in previous work [15] and a new cyclic/cd,fft protocol. Analytic models are developed for every candidate broadcast protocol. The performance metrics of interest are the average server bandwidth, the maximum client delay and the average client delay. The average server bandwidth is defined as the average quantity of data transmitted by the data server in the unit time. The maximum and the average client delay are respectively the longest and average elapse time from the moment the request is sent by the client to the time instant the requested data file is completely received by the client. The analytic models with the single-cell or SFN broadcast transmission schemes, for the average server bandwidth, the average client delay and the maximum client delay, are exact given the model assumptions. The analytic models with the hybrid broadcast transmission scheme, give only approximate results.

Simulation models for the candidate mobile broadcast protocols are developed and used to assess the performance of the protocols as well as the accuracy of the approximate analytic models. The mobile broadcast protocol parameters that have significant impact on performance are varied one at a time. These parameters include the maximum allowable client delay, the server bandwidth used for a single SFN broadcast divided by that used for a single-cell broadcast, the hybrid broadcast threshold, and the number of cells in the broadcast area. From the simulation experiments, it is shown that the hybrid broadcast transmission scheme together with the cyclic/cd,fft protocol provides the best weighted average server bandwidth usage and the

SFN broadcast transmission scheme together with the batching/cbd protocol provides the best average delay performance for a given batching delay parameter and maximum client delay.

1.4 Thesis Organization

The remainder of this thesis is organized as follows:

- Chapter 2 presents background material on LTE, LTE broadcast, its future perspectives and related research studies.
- Chapter 3 reviews the previous scalable on-demand broadcast scheduling protocols that are suited for the single-cell mobile environment.
- Chapter 4 introduces three candidate broadcast transmission schemes and two multi-cell broadcast scheduling protocols for multi-cell on-demand broadcast. Six mobile broadcast protocols are proposed for performance analysis.
- Chapter 5 describes the performance evaluation methodology and presents performance results for the mobile broadcast protocols.
- Chapter 6 gives the thesis summary, presents thesis contributions and discusses the future work.

Chapter 2

LTE Broadcast

This chapter presents an overview of LTE and LTE broadcast as well as the other background information related to this research. Section 2.1 briefly reviews the evolution path of mobile technology from the '1G' technology to the various important LTE releases that have been published in the recent years. Section 2.2 introduces the design for LTE broadcast. Section 2.3 explains the possible use cases in which the LTE broadcast would be useful in the future. Section 2.4 discusses some previous studies on LTE broadcast.

2.1 Overview of LTE

Since the inception of the first generation cellular systems in the early 1980s, mobile telecommunication technology has been evolving rapidly and widespread adoption of a new generation mobile technology has taken place approximately every ten years. In the early 1990s, the first digital mobile technology was introduced in the mobile market as the '2G' (Second Generation) technology, which was the replacement for the preceding '1G' analog technology. The '2G' technology brought about popular mobile data services such as the Short Message Service and the Multimedia Messaging Service. The radio bandwidth spectrum in the mobile network started to become the bearer for both data traffic and voice traffic. In the early 2000s, the 3G mobile broadband communication technology was gradually deployed and enabled around the world. With increased data transmission bit rate, the '3G' technology provided mobile broadband access for the mobile users to receive the data service with improved user experience. The mobile data service began to rival the wired connection service and other wireless connection service.

In the past ten years, sales of mobile devices like smartphones, mobile PCs and tablets have led to large increases in the use of data-oriented mobile applications. In mobile networks, mobile data traffic has far exceeded mobile voice traffic. In anticipation for higher mobile data rates, in 2008 the 3rd Generation Partnership Project (3GPP) started the on-going development of standards for LTE, which was intended to be the next-generation (4G) mobile communication technology [18]. The LTE release 10, also known as LTE-Advanced, was finalized in 2011 and generally regarded as a developmental milestone on the evolution path of LTE, for it was the first 4G standard that met all requirements of the IMT-Advanced standard for wider bandwidths and improved spectrum efficiency [38]. Various enhancements had been incorporated in the LTE 4G standards, including carrier aggregation, enhanced multi-antenna transmission, heterogeneous network deployment, relay node deployment, and CoMP (Coordinated Multipoint) transmission and reception [9, 38]. LTE 4G is capable of providing enhanced mobile data solutions with more flexible use of the radio frequency bands, higher transmission bit rates and lower cost for the high quality data service on the common core network, not only for terminal access but also for wireless backhauling [8, 51]. The notable enhancement particularly related to this thesis is eMBMS (evolved Multimedia Broadcast/Multicast Service), which was inherently supported by LTE 4G.

2.2 Background of LTE Broadcast

The MBMS (Multimedia Broadcast/Multicast Service) was first defined in 3GPP Release 6 in 2004 [23]. Prior to MBMS, the mobile broadcast service in UMTS (Universal Mobile Telecommunications System) had to rely on the point-to-point connection of unicast transmission. With MBMS, a mobile network is able to support not only point-to-point unicast transmission but also point-to-multipoint transmission for mobile broadcast service [22]. Compared to unicast, the point-to-multipoint broadcast design of MBMS enables a mobile network to make more efficient use of radio bandwidth for delivering the same data content to a large number of clients. In a 3G mobile network with broadcasting capacity, a central node referred to as the RNC (Radio Network Controller) is required by MBMS to initiate and synchronize the point-to-multipoint transmission within all its subordinate cells [22]. The use case of MBMS is mainly targeted at push-based delivery of data content to a large audience following a fixed schedule, such as mobile TV and live event broadcast [10, 23]. The competing wireless communication technologies to MBMS include DVB-H, which is the digital terrestrial TV broadcast [22].

In the development of LTE, some basic MBMS functionalities were first incorporated in LTE release 9 in 2009. MBMS in LTE was initially redesigned to comply with the flat LTE architecture without the control node RNC [22]. The optimized MBMS in LTE continued to evolve in later LTE releases and is recognized as eMBMS, which is also known as LTE broadcast. LTE 4G supports high bit rate for data transmission, flexible and efficient spectral utilization, and advanced air interface which enables a new transmission scheme called Multimedia Broadcast/Multicast Service over the MBSFN [5, 29]. Based on the 4G mobile network infrastructure, the eMBMS is able to exploit advanced features for data transmission and provide improved broadcast transmission performance. In order to make efficient use of the radio bandwidth, the eMBMS enables two broadcast transmission schemes: the point-to-multipoint single-cell broadcast transmission in which the LTE broadcast does not require scheduling coordination between the adjacent cells and the point-to-multipoint multi-cell transmission in which a logical node is required to coordinate the broadcast transmission over a cluster of contiguous cells [22]. The logical node in LTE 4G networks is called MCE (Multiple-cell Coordination Entity), which is the controller node of the MBSFN area. Similar to the RNC in UMTS, the MCE defines the radio configurations for its subordinate base stations and allocates radio resources for multi-cell transmission [22]. The shared broadcast channels are only accessible in the cells of the same MBSFN area controlled by a common MCE. The broadcasts of data can be synchronized across the mutually exclusive MBSFN areas only if a node is set up for coordinating the different MCE's. 3GPP defines the MCE to be deployed either as a separate physical node or as an integrated part of the base station [22]. Within the MBSFN, the mobile device receiver may accept signals of the same data content from multiple cells with different delays. The mechanisms for handling the multi-path components of the single-cell point-to-point transmission can be effectively adapted for handling the multi-cell transmission signals without incurring inordinate system complexity.

Traditional unicast transmission is capable of distributing data content in response to a wide variety of user demands where every user is requesting a different data content. The main drawback of unicast transmission is that when a large number of outstanding requests are directed at the same data content, unicast transmission may not be as efficient as mobile broadcast. The use of broadcast transmission as a complement to unicast transmission in mobile networks has already been addressed in eMBMS from LTE release 11 [22]. In LTE 4G networks, the base stations collect feedback information from the mobile devices and forward such information to the common MCE. Based on feedback information on users' data requests, the MCE is able to use heuristics to dynamically select the most efficient transmission scheme among available schemes for the data service. The choice for the transmission scheme can be the point-to-point unicast transmission, point-to-multipoint single-cell broadcast transmission or point-to-multipoint multi-cell transmission [22]. Mobile users might be constantly on the move in and out of a cell. The feedback mechanism has to be effective enough for collecting accurate information on users' data requests and simple enough for implementation, which otherwise may lead to undesirable system overheads.

While an optimal solution for selecting the transmission scheme has not been formulated, some heuristics have been proposed which provide a trade-off between the implementation complexity of the radio interface and the feedback information on users' data requests. One approach keeps track of a count of the number of outstanding requests for the same data content in every individual cell [22]. A reasonable threshold, defined as a certain number of outstanding requests in a cell, is used for initiating the switch between the basic broadcast transmission schemes. If the feedback indicates that the number of outstanding requests in a cell has not passed the threshold, then point-to-point unicast transmission is used by default. Otherwise, point-to-multipoint broadcast transmission should be applied in that cell to replace the unicast transmission. If the same data content is requested from all cells in the same MBSFN area and all cells have the point-to-multipoint broadcast transmission in place, then point-to-multipoint multi-cell transmission should be applied instead of single-cell broadcast transmission for large-scale data dissemination to all cells.

One of the issues with mobile transmission is inter-cell radio interference which may degrade signal reception quality at the boundaries between adjacent cells. To resolve this problem, a common approach would be coordinating data transmission of the neighbouring cells by dynamically allocating the complementary parts of the available radio spectrum to adjacent cells. In LTE 4G networks, the cell-edge interference can be resolved by HetNets/Small Cells, CoMP, and SFN broadcast transmission. HetNets/Small Cell

(Heterogeneous Networks using Small Cells), also referred to as the soft cell, introduces complementary low-power base station nodes near the cell edge under the coverage of an existing macro-node layer. The low-power base station nodes are deployed for offloading the data traffic. The combined use of the low-power base station nodes and the macro nodes for the data service reduces the energy consumption and the deployment/operational cost. [8, 18, 21]. CoMP (Coordinated Multi-Point) uses multiple nearby radio access network nodes for serving the same data request from an end-user device at the cell edge [8, 18, 21]. The radio access network nodes which may geographically be located in the different cells are tightly coordinated using CSI (Channel Status Information). With SFN broadcast, the identical signals are tightly synchronized in time for conveying the data content to every cell within the MBSFN area [6, 46]. The inter-cell coordination in the MBSFN area ensures smooth handover for the moving mobile devices. In the LTE 4G network, the potentially destructive inter-cell radio inference at the cell edge can be harnessed as a enhanced source of useful radio signal for data transmission.

2.3 Future Prospects of LTE Broadcast

A study carried out by Ericsson in 2014 shows the traffic growth of data and voice service between 2010 and 2014[1] (See Figure 2.1). In 2010, the volume of mobile data traffic was roughly the same as the volume of voice traffic. During the four years between 2010 and 2014, voice traffic per quarter year remained at a stable level, while mobile data traffic experienced approximately exponential growth with sixty percent per year growth rate. By September of 2014, the traffic volume of mobile data service became eight times greater than that of voice service. Mobile networks now mostly carry data traffic instead of voice service traffic. The Ericsson report noted that almost two-thirds of mobile data traffic comes from smartphone data subscriptions and the remaining one-third comes from mobile PC's, tablets and mobile routers. Also, it is anticipated in the report that mobile data service subscriptions and mobile traffic per active subscription per year will continue to increase in the next six years. If mobile data traffic keeps on increasing at the current growth rate, mobile data traffic volume will have increased 8-fold by the end of 2020.[1]

Classifying mobile data traffic by the media format, the majorithy of today's mobile data traffic is video. Video traffic first exceeded 50% of total mobile data traffic on cellular networks in 2012.[2] Mobile video is forecast to increase 13-fold between 2014 and 2019, accounting for seventy-two percent of total mobile data traffic by the end of 2019.[2] This indicates that requesting videos through mobile radio channels is really the main contributor to data volume in today's mobile network, and this user behaviour will only be reinforced in the foreseeable future.

[1] http://www.ericsson.com/res/docs/2014/ericsson-mobility-report-november-2014.pdf, access 14-July-2016

[2] http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html, access 14-July-2016

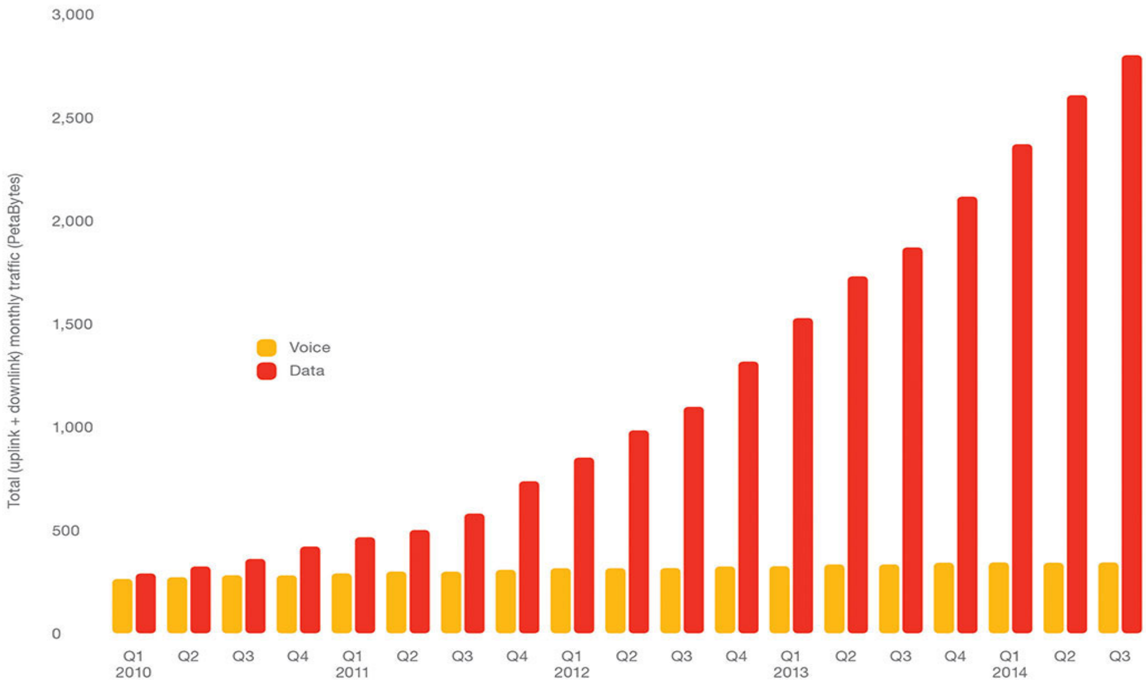


Figure 2.1: The growth in data traffic between Q3 2009 and Q4 2013, Source : Ericsson[1]

In areas like airports, subway stations, and sports arenas, certain popular content such as the flight schedule information, the subway route information, and information concerning the sporting events, tend to have a high likelihood of being requested by multiple mobile users concurrently. Mobile radio bandwidth resources are always extremely valuable in such hot-spot sites. In order to reduce data traffic, LTE broadcast can efficiently distribute some highly popular data content to provide scalable service for mobile data requests, even during peak hours. The use cases for on-demand download broadcast also include the routine upgrade of mobile apps and the delivery of mobile newsletters. For these latter applications, the data transmission should be carried out during off-peak hours through LTE broadcast channels so that the same requested data content is pushed to mobile devices within the same broadcast region with fairly low server bandwidth requirement. For on-demand download broadcast of the mobile data, all requesting clients can share the same broadcast channel to receive the same popular data content. The service quality will not deteriorate regardless of the number of user devices requesting the same data on the same channel. A dedicated high data rate radio channel can be allocated for broadcast transmission to ensure premium quality data service for applications such as distributing high-quality video stream content to a large audience.

With the advent of the eMBMS from LTE-advanced, several LTE broadcast reports have promoted the

[1]http://www.ericsson.com/res/docs/2014/ericsson-mobility-report-november-2014.pdf, access 14-July-2016

use of LTE broadcast as a means of alleviating the mobile data traffic burden.[3,4] In these reports, the possible use cases supported by LTE broadcast include the following,

- Live event streaming,[3,4]
- Real-time TV streaming (mobile TV),[4]
- Video kiosk or video on demand,[4]
- Group information distribution,[3,4]
- Broadcast music and radio,[4]
- Connected car,[4]
- Fixed LTE quadruple play,[4]
- Local area data dissemination (local information such as coupons),[4]
- Stadium wide live event applications,[4]
- Wireless emergency alerts,[4]
- News, stock market reports, weather, and sports updates,[4]
- Firmware/OS updates,[4]
- Off-peak media delivery (e-Newspapers and e-Magazines),[4]
- Data feeds & notifications,[4]
- Pushed video ads,[4]
- Internet of things (smart meters).[4]

2.4 Related Research Studies on LTE Broadcast

SFN broadcast transmission from LTE 4G enables the same data content to be distributed simultaneously to every cell in the same MBSFN area, and this was first introduced in eMBMS. The performance of SFN broadcast transmission had been studied in the past, there were also prior studies on joint delivery of unicast and broadcast in MBSFN networks which demonstrated that the improved user throughput and energy efficiency could be provided by the hybrid approach [16, 17, 34, 45]. Ibrahim *et al.* [26, 27, 28] evaluated the SINR (Signal to Interference Noise Ratio) of MBSFN broadcast transmission and they confirmed the

[3] https://www.qualcomm.com/documents/content-all-potential-lte-broadcastembms-white-paper, access 14-July-2016
[4] http://www.expway.com/wp-content/uploads/White-Paper-14-LTE-Broadcast-Use-Cases-final.pdf, access 14-July-2016

benefits of the constructive cell-edge interference from MBSFN broadcast transmission. Alexiou *et al.* [5, 6] investigated the communication cost of MBSFN in LTE with different network topologies, MBSFN deployments and user distributions. They determined that the number of cells in the MBSFN area would directly affect the performance of MBSFN transmission in terms of total communication cost, and estimated the number of neighbouring rings of cells to be included in the same MBSFN area that would yield the most efficient MBSFN deployment with the lowest possible communication cost. The overall spectral efficiency of the MBSFN broadcast transmission can be maintained even when the size of the MBSFN area is confined to be no more than three neighbouring rings (See Figure 2.2) [5, 6, 40]. In order to handle handover between different MBSFN areas, Nguyen *et al.* [36] proposed a new method to supplement the eMBMS from LTE release 11 by ensuring the service continuity of LTE SFN broadcast transmission for the mobile users while moving across different cells, through different MBSFN areas and on different radio frequencies.

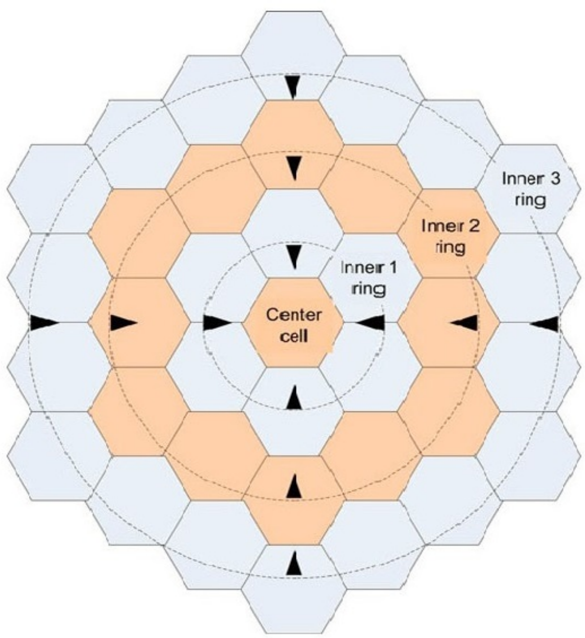


Figure 2.2: A commonly-considered network topology of the MBSFN area [5]

To understand the possible use of LTE broadcast in a real world scenario, Erman *et al.* [19] performed a study on user behaviour and traffic demand of the 2013 Super Bowl attendees. The data collected during the event included both the uplink and the downlink traffic of the mobile data in the stadium. From the collected dataset, they observed that during the Super Bowl event, traffic usage in different areas was not uniform over time. Web content was the major source for mobile traffic both in the downlink and uplink. Video consumption made up a large portion of the downlink traffic, although the number of the video subscribers was fairly small. From the analysis, they suggested that by combining LTE multicast with web content caching, the common requests from users could be served effectively in large events like the Super Bowl. A dynamic scheduling algorithm for resource allocation might also be helpful for dealing with traffic that is not uniform at different sites over time [19].

In order to reduce the traffic demand on the bottleneck access link, Finamore *et al.* [20] proposed a solution by content pre-staging and content caching. They first confirmed that the content downloadable by a mobile terminal is suitable for caching through measuring the popularity, cacheability and object lifetime of a traffic dataset. After implementing the mobile broadcast capacity in the pre-staging system, they showed that the wireless link load can be reduced and data transmission performance from the end-users' perspective can be improved, even in conservative scenarios where cache size is limited and cacheable objects have to be bundled.

2.5 Summary

LTE 4G is the next generation mobile communication technology after 3G; it provides higher data bit rate and improved spectrum efficiency. The mobile broadcast service (eMBMS) has been officially incorporated as part of the LTE 4G specifications. LTE broadcast enables various innovative designs such as the complementary use of unicast and broadcast for the data service, and the SFN broadcast transmission in the defined MBSFN area. By relying on the LTE 4G network infrastructure, mobile broadcast might be able to effectively serve as an alternative approach to the traditional point-to-point unicast transmission for large-scale data dissemination in mobile networks. Based on the growth trend of mobile data in the past, data traffic in mobile networks is expected to increase at a substantial rate over the next few years. LTE broadcast is a possible solution for alleviating the mobile data traffic burden. Some related research work on LTE broadcast has evaluated different aspects of mobile broadcast service.

Chapter 3

On-demand Broadcast Scheduling Protocols

There are two types of broadcast scheduling approaches from previous work: push-based broadcast and pull-based broadcast. In push-based broadcast, historical data access statistics or a set of pre-defined request profiles are assumed as prior knowledge for carrying out the broadcast program scheduling [31]. TV/Radio networks are the typical push-based broadcast systems in which data transmission follows a fixed schedule and data flow is formed only in one direction from the data sender to the data receiver. A pull-based broadcast system is a two-way interactive system in which the data content is transferred based on the outstanding data requests. This is analogous to the classic client/server model. The only difference is that multiple clients are concurrently served in the same broadcast transmission. In the LTE mobile network environment, push-based and pull-based broadcast protocols may be adapted for mobile data service. Their application should account for different use case scenarios. For example, the push-based broadcast protocols are suited for providing mobile TV service. Various pull-based broadcast protocols, such as batching or cyclic scheduling protocols, may be applied for scalable on-demand broadcast in the mobile network environment.

This chapter reviews the on-demand broadcast protocols. Specifically, Section 3.1 examines the on-demand broadcast protocols in three different categories which are on-demand batching, on-demand cyclic and on-demand combined protocols. Section 3.2 investigates an on-demand batching protocol, an on-demand cyclic protocol and an on-demand combined protocol, that are all relating to the mobile broadcast applications in the single-cell environment.

3.1 On-demand Broadcast Scheduling

3.1.1 On-demand Batching Protocols

With on-demand batching protocols, outstanding requests are batched together and server by the same transmission of the data file. The requests that arrive during the broadcast are arranged to be served in the next transmission of the data file. The batching protocols need some rule for deciding which batch of waiting requests to serve next. Some batching protocols that have been proposed for on-demand broadcast include FCFS, MRF, SSTF, SRST, LWF, PLWF and $R\text{x}W$ [2, 3, 49, 50]. With the FCFS (First Come First Serve) protocol, outstanding requests for different data items are served based on their arrival sequence. At the time of scheduling, the data item for broadcast is selected to be that for the request that has been waiting for the longest time [49, 50]. The MRF (Most Requested First) protocol selects the data item for broadcast that has the most pending requests [49]. When choosing between data items with the same number of pending requests, the protocol could either break ties in an arbitrary manner, or in favour of the data item with the lowest measured request probability since it is less likely that more requests for that item will arrive in the near future and join on existing waiting batch [49]. The SSTF (Shortest Service Time First) protocol chooses the requested data item with the shortest service time to broadcast next, after the end of the previous broadcast transmission [2]. When there is a need to break ties, FCFS can be applied [2]. The SRST (Shortest Remaining Service Time) protocol is the preemptive version of the SSTF protocol in which the SRST criterion is applied for selecting the data item for broadcast transmission [2]. A broadcast with longer service time can be interrupted and replaced by a broadcast with shorter service time, and later be resumed from where it was interrupted. With the LWF (Longest Wait First) protocol, the data item selected to be broadcast next is the one for which the total waiting time of all pending requests is largest [2, 49]. The PLWF (Preemptive Long Wait First) protocol incorporates preemption with the LWF rule [2].

Acharya *et al.* [2] showed that for some batching protocols, such as SSTF and LWF, their natural preemptive variants had substantially improved performance in terms of the average response time and the ratio of the response time for a request to its service time for the broadcast transmission. The drawback of the preemptive scheduling design is that the processing overhead and memory requirement increases and more complexity will be added to the scheduling system [2]. Wong *et al.* [49] evaluated the various broadcast batching protocols and observed that LWF scheduling yields the best response time performance even though LWF incurs more scheduling overhead than the other considered protocols such as FCFS and MRF.

The scheduling overhead may thwart the scalability of the broadcast system because the high overhead would limit the number of requests the system can handle. In order to achieve a balanced trade-off between response time performance and the scheduling overhead, Aksoy *et al.* [3] proposed a novel on-demand broadcast scheduling approach, RxW, in which R denotes the number of outstanding requests for the same data item and W denotes the waiting time of the oldest outstanding request for the same data item. RxW was designed to combine the benefits of MRF and FCFS and overcome the high scheduling overhead from LWF. Either the most popular data item or the data item with the oldest outstanding request would have a chance to be selected to be transmitted first. For each requested data item, the arrival time of the oldest outstanding request and the number of the accumulated outstanding requests are recorded in a data structure implemented at the data server. Whenever a new request arrives at the server, the initial request arrival time is created if that data item is requested first, and the R value for that data item is incremented. For each broadcast scheduling decision, the RxW value which is computed from the R value and the waiting time of the oldest outstanding request is updated for all data items, and the data item with the largest RxW value is chosen to be broadcast [3].

Liu *et al.* [31] studied on-demand broadcast scheduling protocols for multi-item requests in the multi-channel broadcast environment. Each client that arrives at the system will issue multiple requests for different data items, and multiple broadcast channels are available for serving equal numbers of requests simultaneously. Liu *et al.* also extended some existing scheduling algorithms designed for single-item requests, like FCFS, to new settings for comparison and analysis [31]. They observed two major reasons that lead to degradation of performance: the request starvation problem and the broadcast mismatch problem. The request starvation problem is that the multi-item scheduling process requires an excessively long time before a requested item is fully delivered. The broadcast mismatch problem is caused by inefficient utilization of multiple broadcast channels. To overcome the performance issues of existing scheduling protocols, Liu *et al.* [32] further proposed a new protocol that quantified the factors for capturing the characteristics of a multi-item, multi-channel request. With this new protocol, the requests are prioritized based on data productivity, which corresponds to the number of outstanding requests pending for the data item and the request urgency. The request urgency is determined by the waiting time for the data item. Through simulation experiments, they showed that the new broadcast scheduling protocol overcame the request starvation problem and the broadcast mismatch problem, and yielded better performance than the other candidate protocols in the multi-item, multi-channel broadcast context.

3.1.2 On-demand Cyclic Protocol

Beside the on-demand batching broadcast protocols, cyclic transmission protocols have also been considered for on-demand broadcast in previous work [7, 11, 12, 13]. For cyclic transmission, the data content needs to be split into a sequence of equal length data blocks. A client should be able to reconstruct the whole file after receiving all the component blocks. The server starts cyclic broadcast transmission by sending out file chunks in order when a client requests the data content. A client making a new request that arrives during a broadcast transmission immediately begins receiving the data blocks from the broadcast channel. The server continues to transmit data blocks, wrapping around to the beginning of the file, as long as at least one client that has requested the file has not received all of the blocks. To support efficient recovery from packet loss, cyclic transmission is sometimes integrated with the erasure coding technique. With erasure coding, the requesting client only receive a certain number of erasure-coded blocks to correctly reconstruct the requested data content, and so, if blocks are lost, clients just keep listening to the broadcast until enough blocks are received [13].

Cyclic transmission can be efficient and reliable when the data request rate is high relative to the data transmission rate. If the clients with outstanding requests have heterogeneous achievable data reception rates, however, the variation on data reception rates becomes a major issue that can significantly impact the cyclic transmission performance. If the data transmission rate is much higher than a client's achievable data reception rate, then that client would experience frequent data loss while receiving the file. If the data transmission rate is considerably lower than a client's achievable data reception rate, then the bandwidth resource of that client would not be fully utilized [11]. To handle heterogeneous clients, the broadcast data server can make use of multiple broadcast channels for transferring the same data content. Each client listens to a subset of all broadcast channels that match up to the client's maximum achievable reception rate [11, 12]. The on-demand data transmission on different broadcast channels may occur simultaneously. The data for the broadcast is divided among the channels and time. At every time instant, the data blocks to be transferred on the channels are different. One potential shortcoming of this multi-channel broadcast design is that a client might receive duplicate data blocks from different channels. By aptly assigning the channels for transferring data content, the number of duplicate data blocks can be reduced or eliminated [11].

3.1.3 On-demand Combined Protocols

Some previous work [15, 48] has also introduced protocols that combine the batching and cyclic transmission approaches for on-demand broadcast. Wolf *et al.* [48] investigated the application of on-demand broadcast for delivering digital products by utilizing the spare bandwidth in a broadcast television delivery system. They imposed restrictions on the broadcast transmission process, in which the number of available broadcast channels was fixed and a deadline time was attached to the content delivery schedule for every subtask [48]. Carlsson *et al.* [15] compared the performance of the batching and cyclic broadcast protocols in terms of the average client delay and the maximum client delay with given server bandwidth. They found that both the batching and the cyclic protocols only provided significantly suboptimal performance over some region of the system design space. The cyclic broadcast transmission protocol provides a maximum client delay close to the best achievable by any protocol when the achievable data reception rate is low relative to the file request rate, while the batching broadcast transmission protocol yields near-optimal performance for the average client delay when the data reception rate is high relative to the file request rate [15]. In order to combine the benefits of both batching and cyclic scheduling protocols, the authors proposed four combined scheduling protocols. For performance evaluation, they evaluated the average or maximum client delay as a function of the average required server bandwidth for the baseline batching, baseline cyclic protocols and the new combined protocols. According to the simulation results, the new combined protocols tend to have better performance than the baseline protocols.

3.2 Protocols for Single-cell Mobile Broadcast and Analysis

With mobile broadcast, data transmission can be carried out not only in every cell but also across the whole radio coverage area. Therefore, a suitable broadcast scheduling protocol for mobile networks should be adaptable for both singe-cell broadcast and multi-cell broadcast, and capable of a smooth transition between those two broadcast modes. This section provides a detailed review of three broadcast scheduling protocols from previous work that can be used within a single cell. The next chapter considers the multi-cell context.

The protocols are described focusing on delivery of a single file that clients are requesting and assumes that the system is able to allocate bandwidth resources for transmission of the file when need. It is, however, desired to achieve a low average bandwidth consumption, by serving many client requests with the same

Symbol	Definition
L	Size of the broadcast data file
λ	Data request rate
$B_{single-cell}$	Average server bandwidth for the single-cell broadcast transmission
B_{SFN}	Average server bandwidth for the SFN broadcast transmission
g	Quotient of the per-cell cost of an SFN broadcast divided by the cost of a single-cell broadcast
D	Maximum client delay
A	Average client delay
r	Data transmission rate
Δ	Batching delay time
T	Hybrid broadcast threshold
N	Number of cells in the mobile broadcast network

Table 3.1: System model notations

broadcast transmission.

3.2.1 Single-cell Batching Protocol and Cyclic Protocol

The batching protocol considered [15] requires a batching delay before data transmission, in expectation that additional client requests may be made for the file that could then be served by the same transmission. With the batching/cbd (batching with constant batching delay) protocol, the batching delay has a fixed length . When a new request is made for the file and no data transmission has been scheduled that has not already begun, a new transmission is scheduled for a time given by the current time plus the batching delay. The data requests that arrive during the batching delay should wait until the beginning of the next scheduled broadcast of data. Using the notation in Table 3.1, with a batching delay time of Δ, the maximum client delay from when a request is made until the data is received is given by $\Delta + L/r$. The operation of the batching/cbd protocol, and that of the cyclic/listeners protocol to be discussed next, is illustrated in Figure 3.1. The bottom arrows in this figure show the arrival times of the client requests, while the arrows at the top show when these requests are completely serviced.

In the model used here of a single cell system, the random arrivals of the data requests follow a Poisson process and the data request rate is denoted as λ. The average time it takes for a new data request to be the first to arrive and cause the next broadcast transmission to be scheduled is $1/\lambda$. With the single-cell batching/cbd protocol, the total average time between transmissions of the data file is equal to the batching delay time Δ plus $1/\lambda$. Thus, the average server bandwidth usage [15] is

$$B_{b/cbd} = \frac{L}{\Delta + 1/\lambda}\,. \tag{3.1}$$

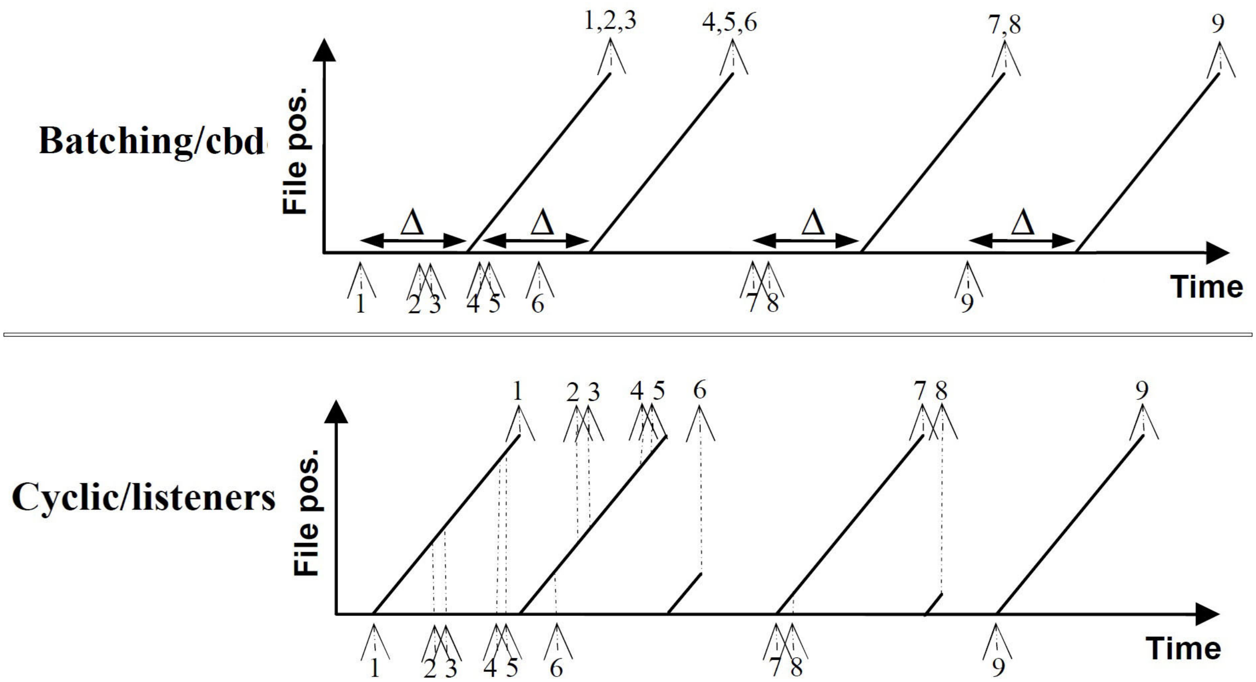


Figure 3.1: Batching/cbd protocol and cyclic/listeners protocol [15]

Before each transmission of the data file, the data request that has arrived first should wait for the batching delay time Δ. Then all data requests that arrive subsequently before the data broadcasting begins will wait for a time period less than Δ. The average number of these randomly arrived data requests is equal to $\lambda\Delta$. The average waiting time for these data requests, before the transmission begins, is $\Delta/2$. Thus, the average client delay and the maximum client delay [15] are

$$A_{b/cbd} = \frac{\Delta + (\lambda\Delta)\Delta/2}{1+\lambda\Delta} + L/r\ ; \tag{3.2}$$

$$D_{b/cbd} = \Delta + L/r\ . \tag{3.3}$$

In the broadcast system, the client delay and the server bandwidth requirement are important metrics concerning the performance of the data service. The average client delay defines the average of the delay time from when the data request has arrived at the data server until the data content has been fully received by the requesting client. The maximum client delay defines the longest possible delay time that might be incurred by any requesting client. The server bandwidth requirement defines the average bandwidth usage required for the data service. To improve the users' experience of the data service, the broadcast system should minimize the average and maximum client delay for data transmission. To optimize the use of limited bandwidth resource, the broadcast system needs to reduce the server bandwidth requirement as low as possible for the same data service. In the same broadcast system, the decrease of the average and maximum client delay should require more server bandwidth to be allocated for delivering the same data content. With batching/cbd, this is accomplished by decreasing Δ. Inversely, a direct reduction in the server bandwidth, with batching/cbd by increasing Δ, would result in increased average and maximum client delay. Therefore, the broadcast system needs to provide a trade-off between the client delay and the server bandwidth requirement. For different types of data service, the clients' sensitivity to the delay time would vary greatly. In media broadcast streaming, the clients are very sensitive to the data transmission delays, and the client-perceived quality of data service is dependent on the delay performance. In an on-demand download service, the clients have higher tolerance regarding the delay, and the client-perceived quality of data service may not be affected as much by the delay performance.

The cyclic/listeners protocol is a baseline cyclic broadcast protocol with a simple design [15, 48]. With cyclic broadcast, a new transmission of the data file would start immediately without delay when a data request arrives with no on-going broadcast of data. When a data request arrives during an on-going broadcast of data, the client would immediately start receiving the data content until the data file is fully delivered. The transmission of the data content is continuously repeating as long as there is at least one requesting client that has not received the complete data file [15]. The data transmission rate for the cyclic protocol is a fixed value, r, regardless the incoming data requests. Assuming that requests arrive according to a Poisson process, for a given time period between $[T\text{-}L/r,\ T]$, the probability that at least one request arrives during that time interval is always 1-$e^{-\lambda L/r}$, because of the memoryless property of the Poisson process. Therefore, the probability that a cyclic transmission of the data file is ongoing at a random time instant T is 1-$e^{-\lambda L/r}$. Thus, the average server bandwidth for the cyclic/listeners protocol is

$$B_{c/l} = r(1 - e^{-\lambda L/r}) \ . \tag{3.4}$$

With cyclic broadcast, there is no waiting time for the data requests. The time it takes for any client to receive the complete data file is equal to the file size divided by the data transmission rate. Thus, the average client delay and the maximum client delay for the cyclic/listeners protocol are

$$A_{c/l} = L/r \ , \tag{3.5}$$

$$D_{c/l} = L/r \ . \tag{3.6}$$

A major difference between the batching protocol and the cyclic protocol is that the change of data transmission rate has a direct impact on the server bandwidth usage of the cyclic protocol, but does not affect that of the batching protocol. For analysis, the data transmission rate for every data request is assumed to be always the same as the client's achievable data reception rate. Given that the requesting clients may each have a different data reception rate, then with the batching or cyclic protocols the data server may utilize multiple broadcast channels of different bandwidths for serving the heterogeneous clients. With just a single broadcast channel, because of the mismatch of the available channel bandwidth and the various data reception rates, these protocols could be prone to performance degradation for the relatively high-bandwidth or low-bandwidth clients. With multiple broadcast channels, the clients with different bandwidths may subscribe to different numbers of broadcast channels concurrently based on their data reception rates for receiving the data content. To reduce or eliminate redundant data transmissions, the requested data blocks of the same file can be arranged to be in different orders in different broadcast channels, or different erasure coded blocks can be transmitted on the different broadcast channels.

Using the batching protocol can be more efficient than using the cyclic protocol when the rate at which the data requests arrive is low relative to the data transmission rate. The reason for this is that when request arrivals are infrequent, the data server with the batching protocol is able to serve more data requests in a single broadcast transmission than the data server with the cyclic protocol, if Δ is chosen to be relatively large. Conversely, using the cyclic protocol would yield better performance than using the batching protocol when the data request rates are high enough for cyclic broadcast transmission to be continuous.

Batching and cyclic protocols are suited for data service in different use case scenarios. The batching protocol ensures in-order delivery of the data content but introduces a batching delay to the data requests. The cyclic protocol does not support in-order data delivery but guarantees immediate transmission for every data request which reduces the client delay time. Accordingly, the cyclic protocol is suited for on-demand broadcast download, while the batching protocol supports both on-demand broadcast download and on-demand audio/video streaming broadcast.

3.2.2 Single-cell Combined Protocol : Cyclic/cd,bot

Protocols that combine batching and cyclic broadcast transmission were proposed in previous work [15]. A performance comparison has been carried out between these combined protocols and the baseline batching and cyclic protocols. In some cases, the combined protocols were shown to be able to achieve near-optimal average or maximum client delay.

The cyclic/cd,bot (cyclic with constant delay bounded on-time) is a combined protocol that provides great performance yet with a simple design. With cyclic/cd,bot, each broadcast transmission starts from the beginning of the file and continues as long as there is at least one request that has not been completely serviced, or the end of the file is reached. A fixed-length batching delay time is imposed before every broadcast transmission begins. When a new data request arrives at a time with no concurrent broadcast transmission, the cyclic/cd,bot broadcast transmission would begin after the constant batching delay time Δ. When a data request arrives during the broadcast transmission, the client would immediately start receiving the data content from the current broadcast transmission until the transmission ends. Then after the batching delay time Δ, the next broadcast transmission begins for serving all unfinished data requests as well as any new requests that arrived during the batching delay time. Like the baseline cyclic protocol, the cyclic/cd,bot protocol is unable to support in-order data delivery but it improves the utilization of the bandwidth resource especially when the data request rate is low relative to the data reception rate. In contrast to the baseline batching protocol, cyclic/cd,bot broadcast does not always deliver the full file during a transmission and at most one broadcast can be in progress at once. Using the notation in Table 3.1, the maximum client delay of cyclic/cd,bot broadcast is $\Delta + L/r$. An approximate analysis for the server bandwidth usage was carried out for the cyclic/cd,bot protocol. It was shown that the approximate analysis had high accuracy since analytic results derived from analysis matched the simulation results. The cyclic/cd,bot protocol turned out to have close to optimal performance in all test cases and it provided an improved trade-off between maximum delay and bandwidth usage [15].

3.3 Summary

This chapter reviewed some on-demand broadcast scheduling protocols from previous work including batching, cyclic and the combined protocols. Among the previous protocols, batching/cbd, cyclic/listeners and cyclic/cd,bot protocols can be adapted for single-cell mobile broadcast. A brief analysis of the single-cell mobile broadcast protocols has been presented. The batching/cbd protocol is suitable for in-order delivery of the data content and provides optimal performance when the data request rate is low relative to the achievable data reception rate. The cyclic/listeners protocol minimizes the delay time for broadcast transmission and provides near-optimal performance when the data request rate is high relative to the data reception rate. The cyclic/cd,bot protocol can combine the benefits of both batching and cyclic protocols.

Chapter 4

Mobile Broadcast Protocols and Analysis

With the development of next-generation mobile technology, the current LTE 4G network is able to support mobile broadcast for data service. The LTE 4G network provides the infrastructure for synchronization of inter-cell broadcast transmission and reduction of cell-edge radio interference. These new features would increase the efficiency of data transmission for mobile broadcast. Aside from the hardware upgrade, another challenge with mobile broadcast is how to come up with an efficient broadcast scheduling design for the multi-cell environment. In previous work, there is a lack of study on multi-cell broadcast protocols. In LTE 4G networks, the mobile broadcast network consists of different contiguous cells. Each individual cell may act independently for serving the data requests in its cell. This is referred to as single-cell broadcast. Additionally, all cells could work together for carrying out broadcast transmissions in the whole network area. This is SFN broadcast transmission, which is enabled by LTE 4G. The existing broadcast protocols have only been studied for single-cell broadcast. Protocols that use both single-cell broadcast and SFN broadcast transmission should have applicable use in LTE 4G networks.

This chapter presents the mobile broadcast protocols for on-demand broadcast data service in the multi-cell mobile environment. Section 4.1 describes the candidate multi-cell broadcast transmission schemes and scheduling protocols, that are all enabled in LTE 4G. Section 4.2 proposes the six mobile broadcast protocols which are constructed from the three broadcast transmission schemes and the two scheduling protocols. The analytic models are developed for every broadcast protocol. The bandwidth performance and the delay performance of the analytic models are evaluated.

4.1 Multi-cell Broadcast Transmission Schemes and Scheduling Protocols

4.1.1 Candidate Transmission Schemes for Mobile broadcast

From a mobile technology like LTE 4G, mobile broadcast might be implemented as a complement to unicast transmission for large-scale data dissemination. In the case when a large number of outstanding requests within the same network are received for the same data content, it would be much more cost-efficient to use the broadcast approach instead of unicast transmission. In this research, three mobile broadcast transmission schemes are proposed, motivated by LTE 4G technologies, namely SFN broadcast transmission, single-cell broadcast transmission, and hybrid broadcast transmission.

In LTE 4G networks, the MBSFN can coordinate inter-cell radio interference and synchronize data transmission in every individual cell within the same MBSFN area. The base station in each cell receives data requests and provides the data. For SFN broadcast transmission, the same data content is transferred to requesting clients in different cells simultaneously as if the whole multi-cell area was regarded as a single cell [22]. Single-cell broadcast was first introduced in 3G network standards and is also enabled in LTE 4G. With single-cell broadcast transmission, data broadcast would take place independently in each individual cell and the data transmissions in different cells are not synchronized. SFN broadcast transmission is suited for the data service if the requests for the same data content arrive frequently and are widely dispersed among all of the network cells. In contrast, single-cell broadcast transmission is suited for the data service if the requests for the same data content arrive frequently and are mostly concentrated in only a few cells.

Since both SFN broadcast transmission and single-cell broadcast transmission can be applied interchangeably on the same network infrastructure in LTE 4G networks, a new hybrid broadcast transmission scheme, which combines the basic transmission schemes, is proposed. For hybrid broadcast transmission, the data server periodically receives feedback information on the data requests from every cellular base station. Given that the feedback information is sufficiently prompt and accurate, then based on the feedback information the data server is able to adaptively select between SFN broadcast transmission and single-cell broadcast transmission for delivering the requested data, according to which would be most efficient. A threshold that defines a certain percentage of cells that have data requests is used in the decision-making process. If the detected percentage of cells with data requests is below the threshold, then the single-cell broadcast transmission is to be applied. Otherwise, the SFN broadcast transmission is used instead. Compared to the two basic transmission schemes, hybrid broadcast transmission is more suited for the broadcast system under changing network conditions with dynamically varying request arrival rates.

4.1.2 Candidate Multi-cell Broadcast Scheduling Protocols

For mobile broadcast, the requested data file is divided into equal-sized data blocks for broadcast transmission. With on-demand data download service, the clients may receive data blocks out of their original order, and store them in a client-side buffer. Once all data blocks are received, the whole data file can be reconstructed in its original order. With an efficient broadcast scheduling protocol, the broadcast system is able to provide the data service with minimized required server bandwidth and guaranteed maximum client delay.

In the multi-cell mobile environment, the broadcast scheduling protocol has to be compatible with the broadcast transmission schemes. Based on the single-cell broadcast scheduling protocols described in Chapter 3, two candidate broadcast scheduling protocols for the multi-cell context are proposed, specifically the batching/cbd protocol and the cyclic/cd,fft protocol.

The batching/cbd protocol as described in Chapter 3 can be implemented in different cells with single-cell broadcast transmission or in the MBSFN area with SFN broadcast transmission. It can also be used with the new hybrid broadcast transmission scheme. Unlike the batching protocol, the basic cyclic protocol is not appropriate for multi-cell content delivery with the hybrid broadcast transmission scheme. With the hybrid broadcast transmission scheme, it would be difficult to synchronize the on-going cyclic protocol data transmissions in different cells so that switching from single-cell to SFN broadcast transmission is feasible. The cyclic/cd,bot protocol would also be more complex to implement with the hybrid broadcast transmission scheme, because of its partial file transmissions.

A modified variant of the cyclic/cd,bot protocol referred to as the cyclic/cd,fft protocol is proposed here for the multi-cell context. With the cyclic/cd,fft protocol, a constant delay is introduced before the beginning of every transmission, which is of the complete data file. The first requesting client that arrives with no on-going or scheduled broadcast transmission would initiate the broadcast transmission after the constant delay. The requesting clients that arrive during the constant delay would wait till the next broadcast transmission. The requesting clients that arrive during the broadcast transmission would immediately start receiving the data content until the end of the current broadcast transmission, then wait for the constant batching delay time, and finish receiving the rest of the data content in the next broadcast transmission. The operation of the cyclic/cd,fft protocol in comparison to that of the cyclic/cd,bot protocol is illustrated in Figure 4.1. The key difference is that the cyclic/cd,fft protocol always makes full-file transmissions, in which the data file is

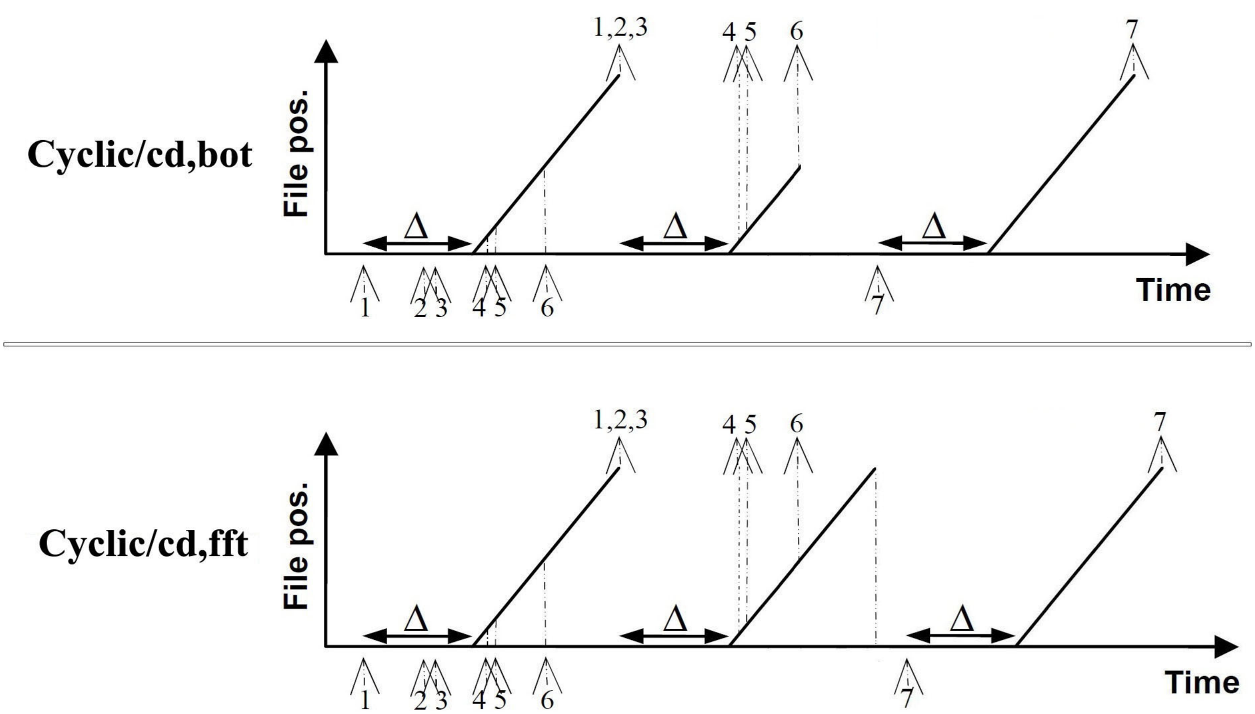


Figure 4.1: cyclic/cd,bot and cyclic/cd,fft protocols in the single cell

transferred in its entirety regardless of the unfinished data requests. This simplifies the design for dynamically switching from single-cell to SFN broadcast transmissions. With both the batching/cbd and cyclic/cd,fft protocols, the length of the batching delay can be made adjustable by the broadcast system in response to the arrivals of the outstanding requests. When the request arrivals are relatively infrequent, the batching delay could be prolonged so that multiple requests would typically be served with each broadcast transmission.

The performance analysis for the batching/cbd protocol in the single cell context is presented in the previous chapter. The average server bandwidth usage with cyclic/cd,fft protocol in the single cell context can be derived as follows. The random arrivals of data requests are, as previously, modelled as following a Poisson process, and the data request rate is denoted as λ. With the cyclic/cd,fft protocol, the data broadcasting may be initiated in two cases. The first case is when no new data request arrives during the previous data transmission. Using the notation from Table 3.1, the probability for this to happen is $e^{-\lambda L/r}$. The average time from the end of the previous transmission until a subsequent request arrives is $1/\lambda$. The average total time from the end of the previous data transmission to the end of the next transmission in this first case is $1/\lambda + \Delta + L/r$. The second case is the complement of the first case, that is at least one new data request arrives during the previous data transmission, then the next broadcast transmission is begun a time Δ following the end of the previous transmission. The probability for the second case to happen is 1- $e^{-\lambda L/r}$. The total time from the end of the previous transmission to the end of the next transmission in this second case is $\Delta + L/r$. Combining these two cases together, the average total time from the end of one transmission to the end of the next with the cyclic/cd,fft protocol becomes

$(1 - e^{-\lambda L/r}) \cdot (\Delta + L/r) + e^{-\lambda L/r} \cdot (\Delta + L/r + 1/\lambda) = \Delta + L/r + \dfrac{e^{-\lambda L/r}}{\lambda}$. Then the average server bandwidth for cyclic/cd,fft broadcast in the single cell context is given by,

$$B_{c/cd,fft} = \frac{L}{\Delta + L/r + \dfrac{e^{-\lambda L/r}}{\lambda}} . \tag{4.1}$$

The maximum client delay for the cyclic/cd,fft protocol is the same as for the batching/cbd protocol, that is $\Delta + L/r$. So with both candidate broadcast protocols, the quality of the broadcast data service can be guaranteed in terms of the longest possible delay time. To derive the average delay time with the cyclic/cd,fft protocol, the operation of the protocol can be viewed as following a repeating pattern. A instance of the pattern begins with the arrival of a new data request which initiates a broadcast transmission after the batching delay time Δ, and ends with a data transmission during which no request arrives. Between the beginning and end, the fixed batching delay time Δ separates every two adjacent data transmissions. There are three types of data requests to be considered. The data request of the first type is the one that initiates a new instance of the pattern. There is only one such data request for each instance of the pattern. It incurs a delay of Δ until data transmission begins. The data requests of the second type are the ones that arrive during the batching delay time Δ before data transmission. The data requests of the third type are the ones that arrive during the broadcast transmission. With Poisson arrivals of the data requests, the number of data transmissions during each instance of the pattern follows the geometric distribution. The parameter p of the geometric distribution is the probability that no data request arrived during the last broadcast transmission, of duration L/r, that is $e^{-\lambda L/r}$. As a result, the expected number of the broadcast transmissions during an instance of the pattern is $1/p$ or $e^{\lambda L/r}$. During an instance of the pattern, the average number of data requests of the second type is $e^{\lambda L/r}(\Delta\lambda)$ and the average number of data requests of the third type is $e^{\lambda L/r}(\lambda L/r)$. The average waiting time before transmission begins for data requests of the second type is $\Delta/2$. The delay time during which no data is being received for data requests of the third type is Δ. The average sum of the client delays divided by the average total number of data requests during an instance of the pattern gives the average client delay. Thus, the average client delay and the maximum client delay for the cyclic/cd,fft protocol are

$$A_{c/cd,fft} = \frac{\Delta + e^{\lambda L/r}(\Delta\lambda)\Delta/2 + e^{\lambda L/r}(\lambda L/r)\Delta}{1 + e^{\lambda L/r}(\Delta + L/r)\lambda} + L/r \ ; \qquad (4.2)$$

$$D_{c/cd,fft} = \Delta + L/r \ . \qquad (4.3)$$

4.1.3 Implementing the Hybrid Transmission Scheme

With batching/cbd, implementing the hybrid transmission scheme is straightforward. In a multi-cell network area, whenever the number of cells with waiting data requests reaches a threshold value, an SFN transmission is scheduled. The start time of this SFN transmission is set to be equal to the earliest start time among the single cell transmissions that had been scheduled for these requests. All the pending requests are served by the SFN transmission, and the scheduled single-cell transmission are cancelled.

With cyclic/cd,fft, the hybrid transmission scheme is implemented in a similar manner. Again, whenever the number of cells for which new single-cell transmissions have been scheduled reaches a threshold value, an SFN transmission is scheduled. The start time of this SFN transmission is set to be equal to the earliest start time among these single-cell transmissions. All requests that would have been served by the scheduled single-cell transmissions are served by the SFN transmission, and the scheduled single-cell transmissions are cancelled.

A complication with cyclic/cd,fft concerns those requests that arrived during a single-cell transmission and are being served by it, but then the next single-cell transmission in their cell (which would be delivering the content they missed from the beginning of the file) is cancelled and replaced by an SFN transmission. If the SFN transmission starts before the end of the current single-cell transmission, the clients will have to switch to receiving the SFN transmission. If such a request had arrived just after the beginning of the single-cell transmission, and the SFN transmission started just before the end of the single-cell transmission, the client delay would be close to twice L/r, since the client would have to listen to the SFN transmission until the end to get the last portion of the file. As long as Δ is at least L/r, however, the maximum client delay would still be bounded by $\Delta + L/r$.

For values of Δ less than L/r, the maximum client delay could still be bounded by $\Delta + L/r$ if the file is erasure coded. Specifically, an erasure coding scheme could be used in which the SFN transmissions deliver erasure coded file blocks different from those delivered in the single-cell transmissions, and such that only an amount of data equal to L (from single-cell and/or SFN transmission) is required to reconstruct the entire file.

4.2 Protocols and Models

With the three available broadcast transmission schemes enabled by LTE 4G and two candidate protocols for multi-cell broadcast scheduling, six mobile broadcast protocols are proposed for on-demand mobile data service. For each of the combinations of broadcast transmission scheme and broadcast scheduling protocol, an analytic model, which is intended to represent the abstract mobile broadcast system, is established for performance evaluation. The performance of the mobile broadcast protocols is measured by the server bandwidth usage, the average client delay and the maximum client delay.

When constructing the analytic models, some assumptions are made for simplifying the problem. The data request arrivals in every cell are modelled as following independent Poisson processes. Thus, the elapsed times between consecutive data requests in each cell are independent random variables following the exponential distribution. The SFN broadcast transmission only takes place in the entire MBSFN area, and is never initiated to just a sub-area. In a real-world broadcast system, the available bandwidth resource is always limited. In the analytic models, the server bandwidth that can be used to support the broadcast service is assumed to be elastic, which means that more bandwidth resources for this service can be allocated when needed. Some factors, such as scheduling overhead and inter-cell radio coordination overhead at the cell edge, are neglected and only the major factors that have the most impact on the overall system performance are included in the analytic models. These major factors are the number of cells N in the MBSFN area, the data request rate λ_i in each cell i ($1 \leqslant i \leqslant N$), the size of the requested data file L, the transmission rate r and the batching delay Δ. Note that, as in the models in Chapter 3, the models focus on delivery of a single data file. With multiple data files, the total bandwidth usage would simply be the sum of that for the individual files. All clients are assumed to be able to receive data at a fixed downlink server transmission rate r.

The server bandwidth usage may differ greatly between SFN broadcast transmission and single-cell broadcast transmission for the same data service. SFN broadcast transmission always serves all outstanding data requests in the MBSFN area. Single-cell broadcast transmission can be applied in certain individual cells and its radio broadcasts are confined within the boundary of each intended cell. The cost of a single SFN broadcast transmission is higher than that of a single-cell broadcast transmission in just one cell, but is expected to have lower cost when compared to making single-cell broadcast transmissions in all cells. This is because with SFN broadcast transmission the data transmissions in different cells are coordinated and the cellular bandwidth resources can be utilized more efficiently. The quotient of the per-cell cost of an SFN broadcast divided by the cost of a single-cell broadcast, which is denoted as g, is used to weight the server bandwidth usage for SFN broadcasts relative to that used for single-cell broadcasts.

In practice, the value of g would be dependent on N. The value of g must lie between $1/N$ and 1. If g was $1/N$, then the total cost of an SFN broadcast transmission for all N cells would be equal to that of a single-cell broadcast in just one cell. SFN broadcast transmissions are most efficient when g is close to this boundary case value. If g was equal to one, then SFN broadcast transmission is not at all useful since there is no difference in terms of per-cell cost between an SFN broadcast transmission and a single-cell broadcast. In a real system, g would be greater than $1/N$ but less than 1.

In the analytic models for protocols using the hybrid broadcast transmission scheme, the value of the hybrid broadcast threshold is an important parameter. The single-cell broadcast transmission is always deployed as the default broadcast approach for data service. Once the number of cells with the common data request reaches the hybrid broadcast threshold, then the SFN broadcast transmission should be applied in place of the single-cell broadcast transmissions. The broadcast scheme would switch back to single-cell broadcast transmissions when the number of cells with the common data request drops below the hybrid broadcast threshold. The threshold value should be an integer between 2 and N. The analytic models developed here could be used to find a near-optimal setting for this value, based on the system parameters.

4.2.1 Batching/cbd with Single-cell Broadcast Transmission

The first proposed mobile broadcast protocol combines the batching/cbd scheduling protocol with the single-cell broadcast transmission scheme. In the mobile network, the batching/cbd protocol is used independently in every individual cell. When a request arrives in a cell with no scheduled broadcast transmission, whether or not there is an on-going broadcast transmission at that time a single-cell broadcast transmission is scheduled to begin in that cell immediately after the batching delay time Δ. The maximum delay for all data requests is bounded by $\Delta + L/r$. The requesting clients that arrive during the delay time Δ would not commence receiving data content until the beginning of the scheduled broadcast transmission.

The average server bandwidth usage in the whole broadcast network is calculated as the aggregate sum of the average server bandwidth usage for all individual cells. To improve the mobile broadcast performance and reduce the server bandwidth usage, the factors that may be adjusted within the broadcast system include the number of cells N, the size of the broadcast data file L and the maximum batching delay Δ. The data request rate in each cell is a factor that comes from the external network environment and can not be controlled by the broadcast system itself. From 3.1, 3.2 and 3.3, the equations for the average total server bandwidth usage, the average client delay, and the maximum client delay are given by

$$B_{b/cbd,single-cell} = \sum_{i=1}^{N} \frac{L}{\Delta + \frac{1}{\lambda_i}} ; \tag{4.4}$$

$$A_{b/cbd,single-cell} = \sum_{i=1}^{N} \frac{\lambda_i}{\sum_{j=1}^{N} \lambda_j} [\frac{\Delta + (\lambda_i \Delta)\Delta/2}{1 + \lambda_i \Delta}] + L/r \tag{4.5}$$

$$D_{b/cbd,single-cell} = \Delta + L/r \ . \qquad (4.6)$$

A state transition model can also be used for analysing the mobile broadcast system. Figure 4.2 presents a continuous-time state transition model for batching/cbd scheduling with the single-cell broadcast transmission scheme. N-cell network is assumed, with equal request rate in each cell. The data request rate in each cell is denoted by λ. The batching delay time is Δ. The state represents the number of cells that have at least one waiting (i.e., not yet receiving data) data request for the same data content. The state space ranges from 0 to N. State transitions only occur between adjacent states. When a transition takes place from state i to state i+1, it indicates that a new data request has arrived in a cell that previously had no waiting data request. The corresponding state transition rate is (N-i)λ. When a transition takes place from state i+1 to state i, it means that a single-cell broadcast transmission has been initiated in a cell and will serve all of its currently waiting requests. The state transition rate from state i+1 to state i is equal to the number of cells that have waiting data requests times $1/\Delta$, that is $(i{+}1){*}1/\Delta$. Given the model assumptions, the considered performance metrics are insensitive to the distribution of the batching delay time, and therefore the rate $1/\Delta$ can be used in the model for the rate at which a broadcast transmission is initiated in a cell, as if the batching delay was exponentially distributed.

Let P_i denote the probability that the broadcast system is in state i with $i \in [0, N]$. To solve this state transition model, the probabilities for all states can be calculated using Algorithm 1. From this state transition model, the average server bandwidth usage, for example, is $L\sum_{i=1}^{N} \frac{iP_i}{\Delta}$. Alternatively, in this case the model could be analytically solved to yield the same equations as 4.4 and 4.5 with all λ_i equal to λ.

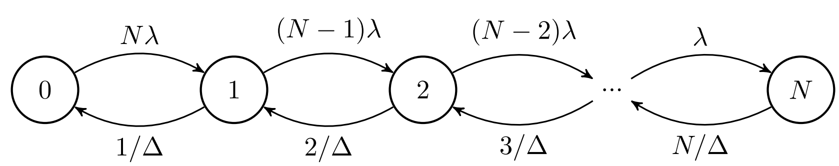


Figure 4.2: The continuous-time state transition model for batching/cbd scheduling with the single-cell broadcast transmission scheme in an N-cell network.

Algorithm 1: An algorithm to calculate the state probabilities in the continuous-time state transition model

Step 1: Under the condition that the rate of transitions into a state must equal that out of the state, for every state, N equations are established for the $N+1$ states.

Step 2: Solve the N equations with $P_0 = 1$.

Step 3: Calculate the sum of all unnormalized probabilities, $P_{sum} = \sum_{i=0}^{N} P_i$

Step 4: Divide the unnormalized probability of each state by the sum, $P_i = \dfrac{P_i}{P_{sum}}$.

4.2.2 Batching/cbd with SFN Transmission

The second proposed mobile broadcast protocol combines the batching/cbd protocol with the SFN broadcast transmission scheme. In the mobile network, the SFN broadcast transmission is used with batching/cbd scheduling across the whole MBSFN area. When a request initially arrives in the MBSFN area with no scheduled broadcast transmission, whether or not there is an on-going broadcast transmission at that time, an SFN broadcast transmission is scheduled to begin in the whole MBSFN area right after the batching delay time Δ. The maximum delay for all data requests is bounded by $\Delta + L/r$. The requesting clients that arrive during the batching delay Δ, would not commence receiving the data content until the beginning of the scheduled broadcast transmission. In this protocol, the broadcast transmission is intended to serve all waiting data requests in the whole MBSFN area. The overall data request rate is the aggregate sum of the data request rates from all cells. Given that the data request rate in cell i is λ_i, the overall data request rate in the MBSFN area is equal to $\sum_{i=1}^{N} \lambda_i$. Given the overall data request rate, the average server bandwidth usage can be directly derived for the whole MBSFN area. The factors that may be adjusted within the broadcast system include the number of cells N, the size of the broadcast data file L and the batching delay Δ. The quotient of the per-cell cost of an SFN broadcast divided by the cost making a single-cell broadcast, which is denoted as g, may also greatly affect the desirability of this protocol. Note that the disparity of data request rates in different cells would have no impact on the average server bandwidth usage. With this protocol, only the total request rate matters. From 3.1, 3.2 and 3.3, the equations for the average server bandwidth usage, as weighted using the parameter g, the average client delay, and the maximum client delay are given by

$$B_{b/cbd,SFN} = \frac{LNg}{\Delta + \dfrac{1}{\sum_{i=1}^{N} \lambda_i}} \; ; \tag{4.7}$$

$$A_{b/cbd,SFN} = \frac{\Delta + \sum_{i=1}^{N} \dfrac{(\lambda_i \Delta)\Delta}{2}}{1 + \sum_{i=1}^{N} \lambda_i \Delta} + L/r \; ; \tag{4.8}$$

$$D_{b/cbd,SFN} = \Delta + L/r \; . \tag{4.9}$$

4.2.3 Cyclic/cd,fft with Single-cell Broadcast Transmission

The third proposed mobile broadcast protocol combines the cyclic/cd,fft scheduling protocol with the single-cell broadcast transmission scheme. In the mobile network, the cyclic/cd,fft protocol is used independently in every individual cell. When a request arrives in a cell with no on-going or scheduled broadcast transmission, then a new broadcast transmission is scheduled to begin in that cell immediately after the batching delay time Δ. The maximum delay for all data requests is bounded by $\Delta + L/r$. The requesting clients that arrive during the batching delay time Δ would not begin receiving the data content until the beginning of the scheduled broadcast transmission. The requesting clients that arrive during an on-going broadcast transmission would immediately commence receiving the data content till the end of the current broadcast transmission. Then after the batching delay time Δ, all unfinished data requests would be served by the next full broadcast transmission.

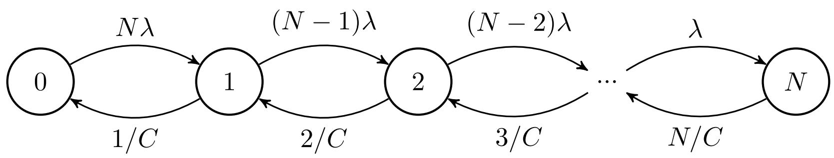
The average server bandwidth usage in the whole broadcast network area, is calculated as the aggregate sum of the average server bandwidth usage for all cells. From 4.1, 4.2 and 4.3, the equations for the average total server bandwidth usage, the average client delay, and the maximum client delay are given by

$$B_{c/cd,fft,single-cell} = \sum_{i=1}^{N} \frac{L}{L/r + \Delta + \dfrac{e^{-\lambda_i L/r}}{\lambda_i}} \; ; \tag{4.10}$$

$$A_{c/cd,fft,single-cell} = \sum_{i=1}^{N} \frac{\lambda_i}{\sum_{j=1}^{N} \lambda_j} \left[\frac{\Delta + e^{\lambda_i L/r}(\lambda_i \Delta)\Delta/2 + e^{\lambda_i L/r}(\lambda_i L/r)\Delta}{1 + e^{\lambda_i L/r}(\Delta + L/r)\lambda_i} \right] + L/r \; ; \tag{4.11}$$

$$D_{c/cd,fft,single-cell} = \Delta + L/r \; . \tag{4.12}$$

As with the combination of batching/cbd scheduling and single-cell broadcast transmission, a continuous-time state transition model can also be used for analysing the mobile broadcast system. Such a model is shown in Figure 4.3, for the case of an N-cell network with equal request rate in each cell. The data request rate in each cell is denoted by λ. The batching delay time is Δ. The state represents the number of cells that have an upcoming scheduled broadcast transmission. The state space ranges from 0 to N. State transitions only occur between adjacent states. When a transition takes place from state i to state i+1, it means that a new data request has arrived in a cell that previously had no upcoming scheduled broadcast transmission. The corresponding state transition rate is $(N\text{-}i)\lambda$. When a transition takes place from state i+1 to state i, it means that a single-cell broadcast transmission has been initiated in a cell. The average time from when a broadcast transmission is first scheduled in a cell, until it begins, is given by C=$L/r + \Delta + (e^{-\lambda L/r} - 1)/\lambda$. This average time can be derived from the probability that the new transmission is scheduled because of an arrival during the previous transmission in that cell, times the average time until that transmission ends plus Δ, plus the probability that the new transmission is scheduled because of an arrival while there is no on-going transmission in that cell, times Δ. Given the model assumptions, the considered performance metrics are insensitive to the distribution of this time, and transition rates can be obtained as if the time was exponentially distributed. To solve this state transition model, the probabilities for all states can be calculated by using Algorithm 1. The average server bandwidth can be calculated as $L\sum_{i=1}^{N-1}\frac{iP_i}{C}$. Alternatively, the model could be analytically solved, to yield the same equations as 4.10 and 4.11 with all λ_i equal to λ.



* $C = L/r + \Delta + (e^{-\lambda L/r} - 1)/\lambda$

Figure 4.3: The continuous-time state transition model for cyclic/cd,fft scheduling with the single-cell broadcast transmission scheme in an N-cell network.

4.2.4 Cyclic/cd,fft with SFN Transmission

The fourth proposed mobile broadcast protocol combines the cyclic/cd,fft protocol with the SFN broadcast transmission scheme. In the mobile network, the SFN broadcast transmission is used with cyclic/cd,fft scheduling across the whole MBSFN area. When a request initially arrives in the MBSFN area with no scheduled or ongoing broadcast transmission, then a new SFN broadcast transmission is scheduled to begin in the MBSFN area after the batching delay time Δ. The maximum delay for all data requests is bounded by $\Delta + L/r$. The requesting clients that arrive during the batching delay time Δ would not begin receiving the data content until the beginning of the scheduled broadcast transmission. The requesting clients that arrive during an on-going broadcast transmission would immediately commence receiving the data content until the end of the current broadcast transmission. Then after the batching delay time Δ, all unfinished data requests would be served by the next full broadcast transmission.

Given that the data request rate in cell i is λ_i, the overall data request rate in the MBSFN area is equal to $\sum_{i=1}^{N}\lambda_i$. Given the overall data request rate, the average server bandwidth usage can be directly calculated for the multi-cell broadcast area using the same reasoning as for equation 4.1. The quotient of the per-cell cost of an SFN broadcast divided by the cost of making a single-cell broadcast, denoted by g, is used to weight the server bandwidth usage. From 4.1, 4.2 and 4.3, the equations for the weighted average server bandwidth usage, the average client delay and the maximum client delay for the combination of cyclic/cd,fft scheduling and the SFN broadcast transmission scheme are given by

$$B_{c/cd,fft,SFN} = \frac{LNg}{L/r + \Delta + \dfrac{e^{-\sum_{i=1}^{N}\lambda_i L/r}}{\sum_{i=1}^{N}\lambda_i}}; \tag{4.13}$$

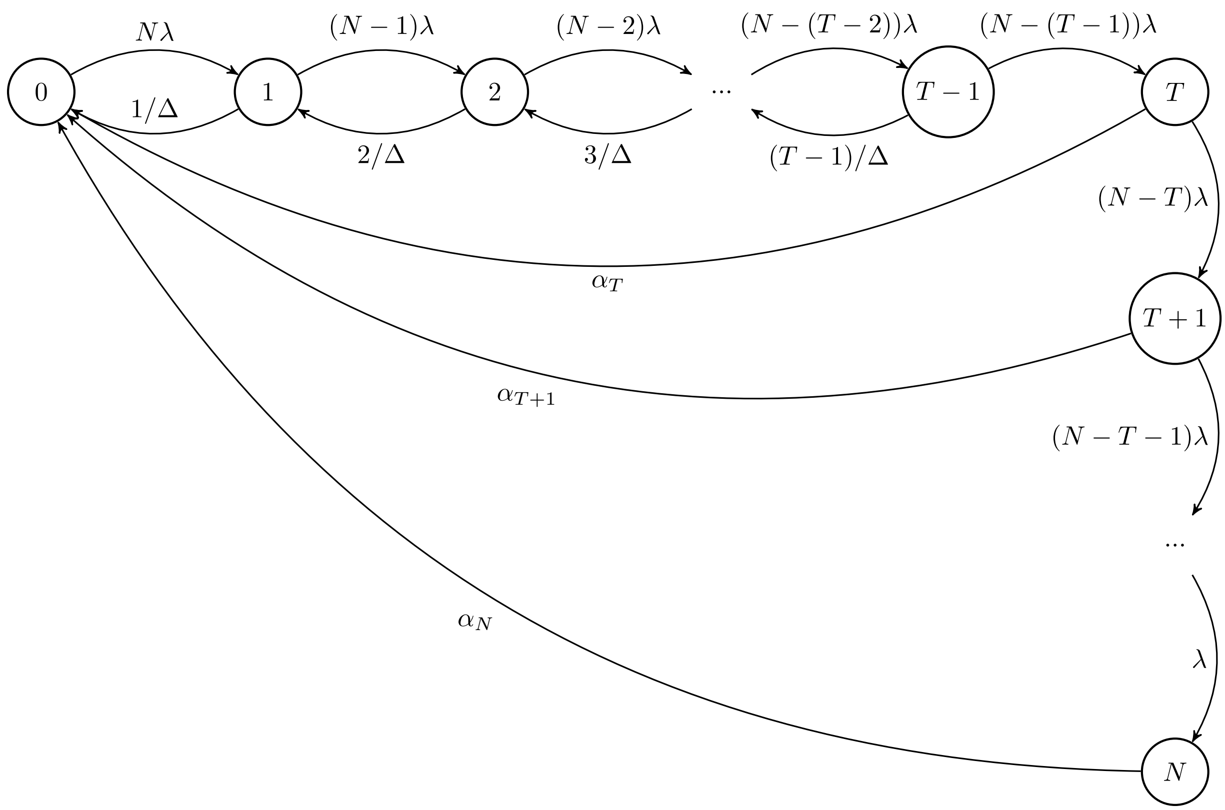
$$A_{c/cd,fft,SFN} = \frac{\Delta + e^{\sum_{i=1}^{N}\lambda_i L/r}(\sum_{i=1}^{N}\lambda_i\Delta)\Delta/2 + e^{\sum_{i=1}^{N}\lambda_i L/r}(\sum_{i=1}^{N}\lambda_i L/r)\Delta}{1 + e^{\sum_{i=1}^{N}\lambda_i L/r}(\Delta + L/r)\sum_{i=1}^{N}\lambda_i} + L/r\ ; \quad (4.14)$$

$$D_{c/cd,fft,,SFN} = \Delta + L/r\ . \quad (4.15)$$

4.2.5 Batching/cbd with Hybrid Broadcast Transmission

The fifth proposed broadcast protocol combines the batching/cbd scheduling protocol with the hybrid broadcast transmission scheme. In the mobile network with N cells, the SFN broadcast transmission and single-cell broadcast transmission may be used interchangeably based on the changing network conditions, together with batching/cbd as the broadcast scheduling protocol. Across the whole MBSFN area, the mobile broadcast server needs to keep track of the number of cells with at least one request for the same data content. If the number of cells with at least one request is equal to or above the threshold T, then the SFN broadcast transmission scheme is used for the whole MBSFN area, otherwise single-cell broadcast transmissions are used. The hybrid broadcast transmission scheme, with appropriate choice of the threshold T, ensures that SFN broadcast transmissions are used only when there are sufficiently frequent arrivals across the MBSFN area so that the bandwidth cost is reduced through the use of SFN broadcast. For the mobile broadcast protocols using the hybrid broadcast transmission scheme, it was possible to devise only approximate models, because of the complexity of the protocols. Continuous-time state transition models are developed for the case of an N-cell network with equal request rate in each cell. The data request rate in each cell is denoted by λ. The batching delay time for any data request is no greater than Δ. Thus, the maximum client delay is bounded by $\Delta + L/r$. Based on the common basic settings, two different asymptotically-exact continuous-time state transition models were developed for each of the hybrid broadcast protocols. Approximations for the weighted average server bandwidth usage and average client delay in the broadcast network can be derived from these models. The models are asymptotically exact for the cases of very low and very high arrival rates.

The first of the continuous-time state transition models for the combination of batching/cbd scheduling and hybrid broadcast transmission is presented in Figure 4.4. The state represents the number of cells with at least one waiting (i.e., not yet receiving data) data request. The state space ranges from 0 to N. For the states 0 through T-1, transitions take place only between adjacent states due to request arrivals and single-cell broadcast transmissions. In state T, the hybrid broadcast transmission threshold is reached and SFN broadcast transmission is used in place of the single-cell broadcast transmission. For states T through N, each state has one transition from the previous state, one transition directed to state 0 (taken when an SFN broadcast occurs) and one transition directed to the next higher-numbered state if it exists. The transition rate from state i to state 0 with $i \in [T, N]$ is an approximate value denoted in Figure 4.4 as α_i. This approximation is based on estimating the average time from when state i is entered, until an SFN broadcast occurs, by Δ (the delay from the arrival time of the first request that will be served by this broadcast) minus



$$* \ \alpha_i = \frac{1}{max[\Delta - (\frac{1}{(N-1)\lambda} + \frac{1}{(N-2)\lambda} + ... + \frac{1}{(N-(i-1))\lambda}), \epsilon]} ; \epsilon = 10^{-6}$$

Figure 4.4: The first continuous-time state transition model for batching/cbd scheduling with the hybrid broadcast transmission scheme in an N-cell network.

the average time taken for the request arrivals in the other i-1 cells that resulted in moving into state i. To solve this state transition model, the probabilities for all state can be calculated using Algorithm 1. Then based on the state probabilities, the weighted average server bandwidth usage is derived using the weighted (using the parameter g) sum of single-cell and SFN broadcast transmission rates. The overall average client delay is derived as the sum of average client delays from two sets of states. One set corresponds to all states in which SFN broadcast transmissions occur, and the other one corresponds to the states with single-cell broadcast transmissions. Specifically,

$$B_{b/cbd,hybrid} = L\sum_{i=1}^{T-1}\frac{iP_i}{\Delta} + LgN\sum_{i=T}^{N}P_i\alpha_i\ ; \tag{4.16}$$

$$A_{b/cbd,hybrid} = (1-\sum_{i=T}^{N}P_i)(\frac{\Delta+(\lambda\Delta)\Delta/2}{1+\lambda\Delta}+L/r) + \sum_{i=T}^{N}P_i(\frac{\Delta+(N\lambda\Delta)\Delta/2}{1+N\lambda\Delta}+L/r). \tag{4.17}$$

The second continuous-time state transition model for the combination of batching/cbd scheduling and hybrid broadcast transmission is presented in Figure 4.5. Again, the state represents the number of cells with at least one waiting data request. The state space is from state 0 to state T-1 plus the combined state $T...N$. From state 0 to state T-1, transitions take place only between adjacent states due to request arrivals and single-cell broadcast transmissions. The state $T...N$ indicates that the number of cells with at least one waiting data request is equal to or greater than the hybrid broadcast threshold. For this combined state, there is only one transition coming from the previous state T-1 and one transition to state 0. The rate of the transition to state 0 is estimated simply as $1/\Delta$. To solve this state transition model, the probabilities for all states can be calculated using Algorithm 1 with N replaced by T. Then based on the state probabilities, the weighted average server bandwidth usage is derived using the weighted sum of single-cell and SFN broadcast transmission rates. The overall average client delay is derived as the sum of average client delays from two set of states, weighted according to the probability of the being in each set. One set corresponds to the combined state $T...N$ and the other one corresponds to all states other than the combined state. Specifically,

$$B^{'}_{b/cbd,hybrid} = L\sum_{i=1}^{T-1}\frac{iP_i}{\Delta} + \frac{LgNP_{T...N}}{\Delta}\ ; \tag{4.18}$$

$$A^{'}_{b/cbd,hybrid} = (1-P_{T...N})(\frac{\Delta+(\lambda\Delta)\Delta/2}{1+\lambda\Delta}+L/r) + P_{T...N}(\frac{\Delta+(N\lambda\Delta)\Delta/2}{1+N\lambda\Delta}+L/r). \tag{4.19}$$

This second model was developed after the first model, to see if a simpler model would give good results. In addition to being simpler than the first model, it is exact for T=1, as well as asymptotically exact for very low and very high request rates as with first model.

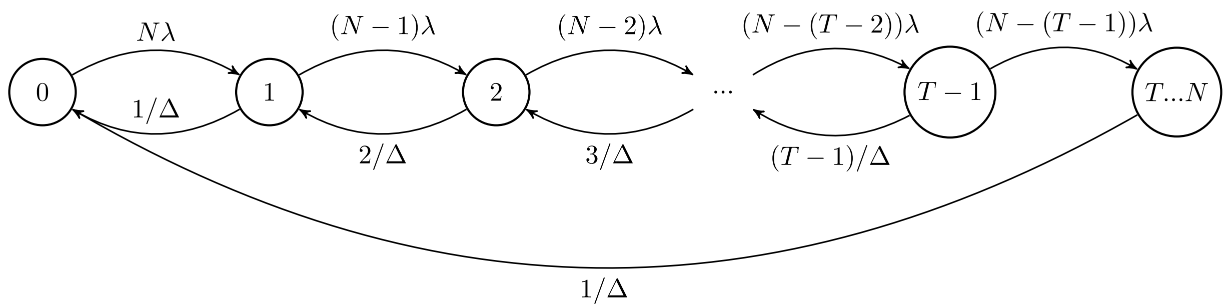


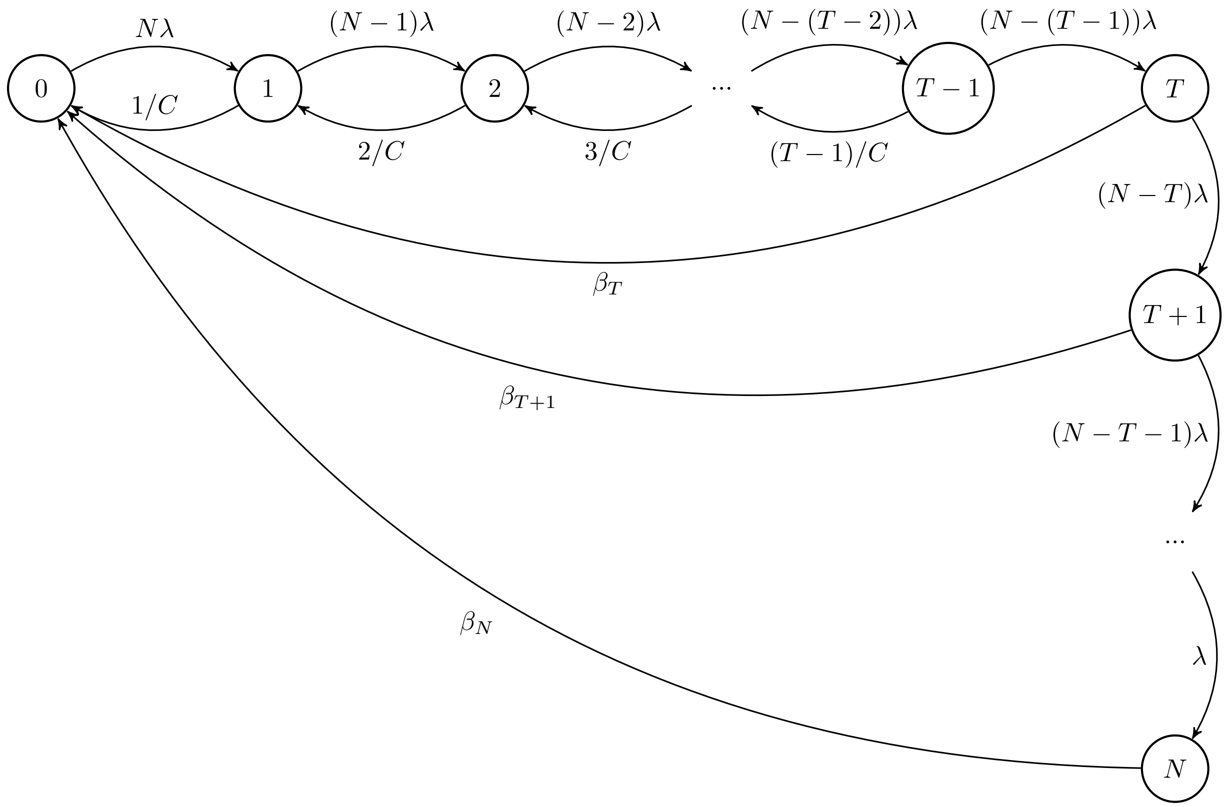
Figure 4.5: The second continuous-time state transition model for batching/cbd scheduling with the hybrid broadcast transmission scheme in an N-cell network.

4.2.6 Cyclic/cd,fft with Hybrid Broadcast Transmission

The sixth proposed mobile broadcast protocol combines the cyclic/cd,fft scheduling protocol with the hybrid broadcast transmission scheme. Like in the previous protocol, the SFN broadcast transmission and single-cell broadcast transmission may be used interchangeably based on the changing network conditions. Across the whole MBSFN area, the mobile broadcast server needs to keep track of the number of cells that have an upcoming scheduled broadcast transmission for the same data content that will serve requests from that cell. If the number of such cells is equal to or above the hybrid broadcast threshold T, then the SFN broadcast transmission scheme is used for the whole MBSFN area, otherwise single-cell broadcast transmissions are used.

Like for the previous protocol using the hybrid broadcast transmission scheme, two different approximate continuous-time state transition models are developed for the case of an N-cell network with equal request rate λ in each cell. The maximum client delay is bounded by $\Delta + L/r$. Approximations for the weighted average server bandwidth usage and average client delay can be derived from the continuous-time state transition models. The models are asymptotically exact for the cases of very low and very high arrival rates.

The first of the continuous-time state transition models for the combination of cyclic/cd,fft scheduling and hybrid broadcast transmission is presented in Figure 4.6. This state transition model is exactly the same as the first continuous-time state transition model for batching/cbd scheduling with hybrid broadcast transmission, except that Δ is replaced by C and α_i is replaced by β_i. C is the same as in the model shown in Figure 4.3, while the transition rates to state 0 from states T, $T \in [1,N]$, denoted by α_i for $i = T$ to N, are estimated in a similar manner as for the corresponding rates in the model in Figure 4.4. As with the previous models for batching/cbd with the hybrid transmission scheme, the weighted average server bandwidth usage is derived using the weighted sum of single-cell and SFN broadcast transmission rates. The overall average client delay is derived as the sum of average client delays from two sets of states, weighted according to the probability of being in each set. One set corresponds to all states in which SFN broadcast transmissions occur, and the other one corresponds to the states with single-cell broadcast transmissions. Specifically,



* $C = L/r + \Delta + (e^{-\lambda L/r} - 1)/\lambda$

$$\beta_i = \frac{1}{max[L/r + \Delta + \dfrac{e^{-N\lambda L/r} - 1}{N\lambda} - (\dfrac{1}{(N-1)\lambda} + \dfrac{1}{(N-2)\lambda} + ... + \dfrac{1}{(N-(i-1))\lambda}), \epsilon]} \; ; \epsilon = 10^{-6}$$

Figure 4.6: The first continuous-time state transition model for cyclic/cd,fft scheduling with the hybrid broadcast transmission scheme in an N-cell network.

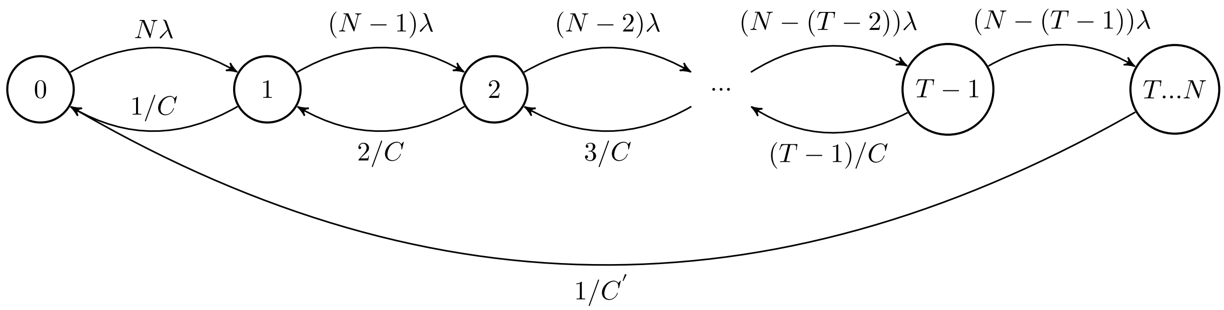
$$B_{c/cd,hybrid} = L\sum_{i=1}^{T-1} \frac{iP_i}{C} + LgN\sum_{i=T}^{N} P_i\beta_i \; ; \tag{4.20}$$

$$\begin{aligned} A_{c/cd,hybrid} = {} & (1 - \sum_{i=T}^{N} P_i)(\frac{\Delta + e^{\lambda L/r}(\lambda\Delta)\Delta/2 + e^{\lambda L/r}(\lambda L/r)\Delta}{1 + e^{\lambda L/r}(\Delta + L/r)\lambda} + L/r) + \\ & \sum_{i=T}^{N} P_i(\frac{\Delta + e^{N\lambda L/r}(N\lambda\Delta)\Delta/2 + e^{N\lambda L/r}(N\lambda L/r)\Delta}{1 + e^{N\lambda L/r}(\Delta + L/r)N\lambda} + L/r). \end{aligned} \tag{4.21}$$

The second continuous-time state transition model for the combination of cyclic/cd,fft scheduling and hybrid broadcast transmission is presented in Figure 4.7. This second state transition model is exactly the same as the second model for batching/cbd with hybrid broadcast transmission with the exception that Δ is replaced by $C^{'}$. By the same calculation process as for the first model, the weighted average server bandwidth usage is derived using the weighted sum of single-cell and SFN broadcast transmission rates. The overall average client delay is derived as the weighted sum of average client delays from two set of states, one corresponds to the combined state $T...N$ and the other one corresponds to all states other than the combined state. Specifically,

$$B^{'}_{b/cbd,hybrid} = L\sum_{i=1}^{T-1}\frac{iP_i}{C} + \frac{LgN(1-\sum_{i=0}^{T-1}P_i)}{C'}; \tag{4.22}$$

$$\begin{aligned} A^{'}_{b/cbd,hybrid} = & \left(\frac{\Delta + e^{\lambda L/r}(\lambda\Delta)\Delta/2 + e^{\lambda L/r}(\lambda L/r)\Delta}{1+e^{\lambda L/r}(\Delta + L/r)\lambda} + L/r\right)\sum_{i=0}^{T-1}P_i + \\ & \left(\frac{\Delta + e^{N\lambda L/r}(N\lambda\Delta)\Delta/2 + e^{N\lambda L/r}(N\lambda L/r)\Delta}{1+e^{N\lambda L/r}(\Delta + L/r)N\lambda} + L/r\right)(1-\sum_{i=0}^{T-1}P_i). \end{aligned} \tag{4.23}$$



* $C = L/r + \Delta + (e^{-\lambda L/r} - 1)/\lambda$; $C^{'} = L/r + \Delta + (e^{-N\lambda L/r} - 1)/(N\lambda)$

Figure 4.7: The second continuous-time state transition model for cyclic/cd,fft scheduling with the hybrid broadcast transmission scheme in an N-cell network.

4.3 Summary

This chapter introduces six mobile broadcast protocols for on-demand data service in the mobile network. These mobile broadcast protocols are proposed from three mobile broadcast transmission schemes and two multi-cell broadcast scheduling protocols. Using some assumptions, such as Poisson request arrivals, analytic models are constructed for performance analysis. The broadcast transmission performance is measured by the average server bandwidth usage, the average client delay and the maximum client delay. With the common parameter settings, the maximum client delay is always bounded by $\Delta + L/r$. For the protocols using the SFN or single-cell broadcast transmission scheme, equations can be directly derived for calculating average bandwidth usage and the average client delay. For the broadcast protocols using the hybrid broadcast transmission scheme, the average bandwidth usage and the average client delay can only be estimated from the approximate continuous-time state transition models.

Chapter 5

Experiments and Results

This chapter presents performance results for the six mobile broadcast protocols from Chapter 4. The main objectives are to determine the performance differences among the protocols, and to assess the accuracy of the approximated analytic models that were developed for the protocols using the hybrid transmission scheme. Simulation programs written in the C language are used for emulating mobile broadcast protocol operation. The data requests randomly arrive in every cell following a Poisson process. To emulate random request arrivals in each cell, the programs employ a random number generating function for creating the required exponentially distributed inter-arrival times. In the simulation experiments, all cells have the same data request rate.

To determine appropriate running time for each simulation run, a given number of contiguous data requests are grouped as a batch. Results for each batch are measured in the simulation according to the operation of the protocol. To ensure that the simulation is statistically valid, the confidence interval level of the collected results is obtained after each batch is processed. Once the confidence interval level is detected to be high enough (e.g. 99%), then the collected results are deemed to be accurate and the final results are computed as the average of results from the batches.

Simulation programs were written for all six mobile broadcast protocols. Since the same assumptions are made in the simulation models as in the analytic models, the simulation results and the results for the exact analytic models are identical except for the very small statistical variation in the simulation results. For consistency, the figures in the following sections present the simulation results even for the protocols using single-cell or SFN transmission, for which exact analytic models have been developed. Section 5.3.1 presents comparisons between the simulation and appropriate analytic model results for the protocols using the hybrid transmission scheme.

5.1 Experimental Plan

The following system and protocol assumption are made. The area for mobile broadcast is N cells, which is regarded as the total size of the MBSFN area. The same data file whose size is L is requested by clients whose requests randomly arrive at the same rate in every cell. For data transmission, all clients are assumed able to receive the file at the fixed data transmission rate r. For SFN or hybrid broadcast transmission, the quotient of the per-cell cost of an SFN broadcast divided the cost of a single-cell broadcast is denoted as g, and assumed to be between $1/N$ and 1. For the hybrid broadcast protocols, the value for the hybrid broadcast threshold is assumed to be a fixed protocol parameter that could be selected based on the size of the MBSFN area and the value of g.

The total time for transferring a complete data file is always L/r. Since the maximum client delay equals $\Delta + L/r$, changing the batching delay Δ is equivalent to changing the maximum client delay D. In the experiments, the maximum client delay D, instead of the batching delay Δ, is treated as an input parameter. The performance metrics measured in the simulation experiments are the average weighted server bandwidth usage and the average client delay. The weighted average server bandwidth usage is derived using the weighted sum of single-cell and SFN broadcast transmission rates. The average client delay is the average elapsed time for any data request starting from the request arrival time until the time instant when the data file is fully delivered. For the protocol combining cyclic/cd,fft scheduling with hybrid transmission scheme, the results for the average client delay assume that the file is erasure coded so that only an amount of data equal to L is required to reconstruct the entire file.

In the first experiment, the six protocols are evaluated under the same default parameter settings, as a function of the per-cell request rate λ. To examine the impact of each other input parameter on broadcast transmission performance, every protocol is then re-evaluated by varying these input parameters one at a time. These parameters include the maximum client delay, the quotient of the per-cell cost of an SFN broadcast divided by the cost of a single-cell broadcast, the hybrid broadcast threshold and the number of cells in the MBSFN area. In these experiments, L and r are, without loss of generality, fixed at 1, equivalent to fixing the unit of data volume to be the file size and the unit of time to be the time required to transfer the file once. For an alternative perspective on performance, further results are shown with the maximum client delay fixed at 1, making it the unit of time, and with different L and r value combinations.

5.2 Results under Default Parameter Settings

Table 5.1 shows the default value and value range for each of the input parameters. Since D=Δ+L/r and $0 \leqslant \Delta < D$, then the maximum client delay D has to be equal to or greater than L/r. The size of the broadcast data file L, that equals $r(D-\Delta)$, could range between 0 and rD. The data transmission rate r, that equals $L/(D-\Delta)$, should be greater than L/D. The mobile network should contain at least 2 cells, that is $N \geqslant 2$. The per-cell cost of an SFN broadcast transmission over N cells should be less than that of a single-cell broadcast, but the total cost across all cells should surely be greater than the cost of a single-cell broadcast in only one cell, so $1/N < g < 1$. One of the main reasons for g being less than 1 is that the strength of SFN transmission signals received at the cell edge may increase when compared with the alternative single-cell broadcast. The increase in the overall transmission performance indicates a reduction in the required server bandwidth usage. The hybrid broadcast threshold should be an integer value between 2 and N.

Without loss of generality, the size of the data file for mobile broadcast is defined to be the unit of data volume, that is L by default is 1. Similarly, the amount of time required for transferring the complete data file is defined to be the unit of time. The data transmission rate r then becomes $L/1$, which equals 1. The default value of the batching delay Δ is chosen to be the same as the data transmission time. Then D by default is $1 + L/r$, which equals 2. For the default parameter settings, the MBSFN area is considered to have 19 cells, which, for example, could be arranged in a round shape with a center cell and two neighbouring cell rings (as in "inner 1 ring and inner 2 ring" of Figure 2.2). The mobile network of this particular shape is a common design considered in previous work [40] since this network design could provide improved spectral efficiency and increased transmission throughput. Some past work suggests that an MBSFN area with 19 cells would be a preferable network deployment when the size of the mobile coverage area is medium [5, 6]. The default quotient of the per-cell cost of an SFN broadcast transmission divided by the cost of a single-cell broadcast is assumed to be the intermediate value 0.5. For the protocol combining batching/cbd with hybrid broadcast, the default value for the hybrid broadcast threshold is set to be 11. For the protocol combining cyclic/cd,fft with hybrid broadcast, the default value for the hybrid broadcast threshold is set to be 8. These two threshold values are chosen since they were found to be the optimal values for these protocols under the default parameter settings.

Parameter	**Value range**	**Default values**
D	$[L/r\ ,\ \infty)$	2
g	$(1/N\ ,\ 1\ )$	0.5
L	$(0\ ,\ rD]$	1
r	$[L/D\ ,\ \infty\ )$	1
$T_{batching/cbd,hybrid}$	$[2\ ,\ N]$	11
$T_{cyclic/cd,fft,hybrid}$	$[2\ ,\ N]$	8
N	$[2\ ,\ \infty)$	19

Table 5.1: The value range and the default values for the test model parameters

Figure 5.1 plots the weighted average server bandwidth usage of each of the six protocols under the default parameter settings. The data request rate per cell is varied from 0.01 to 100. The server bandwidth usage of every protocol steadily increases and slowly stabilizes at a certain value once the request arrivals are frequent enough. An exception is that for the protocol combining cyclic/cd,fft with hybrid broadcast, the average server bandwidth usage has a slight decrease just before reaching the stabilized value, reflecting a transition point between making mostly single-cell transmissions and making mostly SFN transmissions. At fairly low data request rates, the two protocols using SFN broadcast have the same weighted server bandwidth usage and the other protocols also have the same server bandwidth usage which is lower than that when using SFN broadcast. When the data request rate becomes sufficiently high and continues to increase, the average server bandwidth usage of the six protocols level out at three different values: the lowest stabilized server bandwidth usage is attained by the protocol combining cyclic/cd,fft with SFN broadcast, and the protocol combining cyclic/cd,fft with hybrid broadcast; an intermediate server bandwidth usage is attained by the protocol combining batching/cbd with SFN broadcast, the protocol combining batching/cbd with hybrid broadcast, and the protocol combining cyclic/cd,fft with single-cell broadcast; and finally the highest server bandwidth usage is attained by the protocol combining batching/cbd with single-cell broadcast.

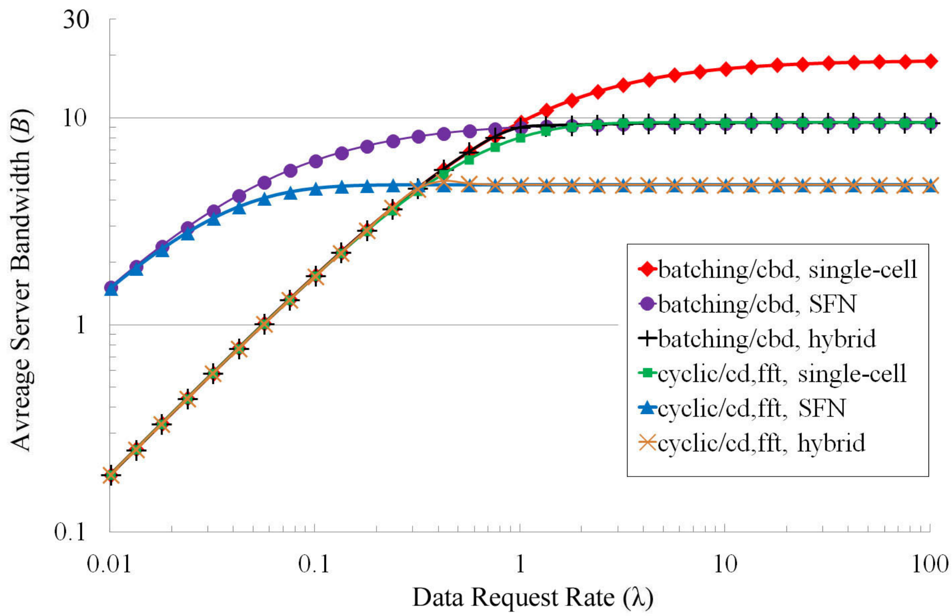


Figure 5.1: The weighted average server bandwidth usage of the six protocols under default parameter settings

Among protocols using the same broadcast scheduling protocol, the weighted server bandwidth using SFN broadcast transmission is higher than that when using single-cell broadcasts if the data request rate is low relative to the data transmission rate. Gradually as the data request rate increases, the weighted average server bandwidth usage when using single-cell broadcasts increases and eventually surpasses that when using SFN broadcast transmissions at a crossover point. The crossover point occurs at a higher data request rate with batching/cbd scheduling than with cyclic/cd,fft scheduling. With both batching/cbd and cyclic/cd,fft scheduling, the weighted server bandwidth usage with the hybrid broadcast transmission scheme is approximately the same for each data request rate as the minimum of the weighted server bandwidth usage with the SFN or single-cell transmission schemes. Specifically, before the crossover point, the hybrid broadcast transmission scheme gives about the same weighted server bandwidth as the single-cell broadcast transmission scheme. After the crossover point, the hybrid broadcast transmission scheme gives about the same weighted server bandwidth usage as the SFN broadcast transmission scheme.

Figure 5.2 presents the average client delay of the six protocols under default parameter settings. The data request rate per cell is varied from 0.0001 to 500. As the data request rate increases within the defined value range, the average client delay with each protocol descends from the same initial value equal to the maximum client delay D and eventually stabilizes at one of the two values. For the protocols with the same transmission scheme, the average delay curves decline at nearly the same rate at fairly low data request rates, with hybrid broadcast transmission scheme and single-cell broadcasts giving the same average client delay, and SFN broadcasts lower average client delay. When the data request rate becomes high enough, the average client delay of the protocols using cyclic/cd,fft scheduling converge at the same value equal to $[\Delta/(\Delta + L/r) \times (\Delta/2) + (L/r)/(\Delta + L/r) \times \Delta]$ plus the file transmission time L/r. The average client delay of the protocols using batching/cbd scheduling stabilize at a lower value equal to one-half of the batching delay Δ, plus the file transmission time L/r.

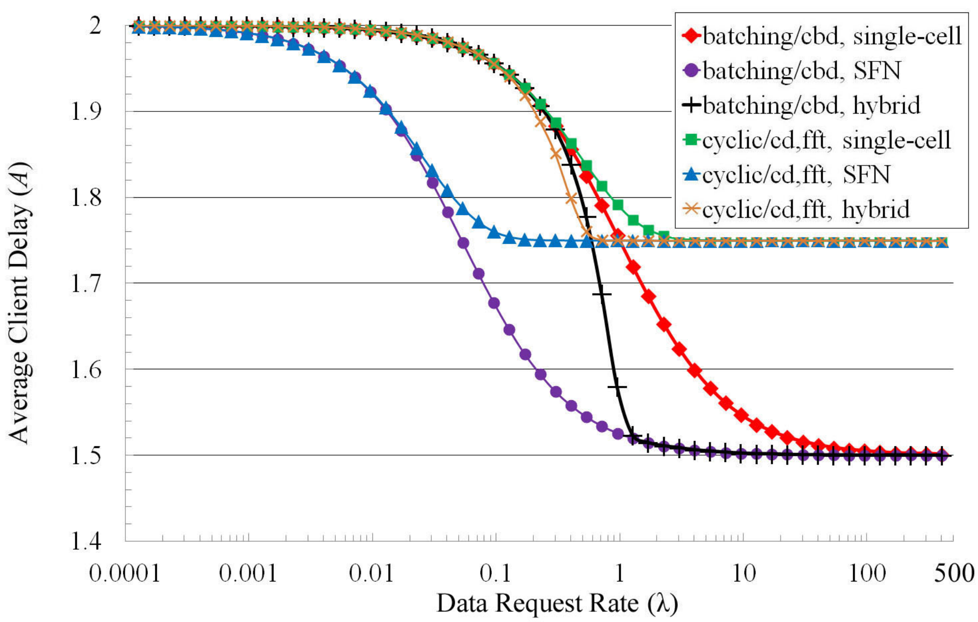


Figure 5.2: The average client delay of the six protocols under default parameter settings

From the performance results for the protocols under default parameter settings, with the same broadcast scheduling protocol using single-cell broadcasts would incur lower cost than using SFN broadcasts when the the data request rate is low relative to the data transmission rate. Otherwise, when the data request rate is high relative to the data transmission rate, using single-cell broadcasts would incur higher cost than using SFN broadcasts. After the data request rate becomes sufficiently high, the weighted server bandwidth usage and the average client delay level off, and broadcast transmissions occur at a regular spacing. Considering the entire range of data request rates, the hybrid broadcast transmission scheme is shown to give the best server bandwidth performance, for each scheduling protocol. The cyclic/cd,fft scheduling protocol yields better server bandwidth performance than the batching/cbd scheduling protocol when the data request rate is high relative to the data transmission rate. As for the average delay performance, SFN broadcast yields the lowest average client delay and the single-cell broadcast scheme gives the highest average client delay. The hybrid broadcast scheme gives intermediate average delay performance. Therefore, under default parameter settings the broadcast protocol with cyclic/cd,fft scheduling and hybrid broadcast transmission has the best server bandwidth performance, and the broadcast protocol with batching/cbd scheduling and SFN broadcast transmission has the best average delay performance. Note however that this comparison is for equal maximum client delay. For equal weighted server bandwidth usage, cyclic/cd,fft scheduling with hybrid broadcast transmission would give lower average client delay than batching/cbd scheduling with SFN transmission.

5.3 Results with Variable *D*

In the mobile broadcast protocols, the data server is able to adjust the maximum client delay by changing the duration of the batching delay. In this section, results are shown for default parameter settings (as in Table 5.1) except for different values of the maximum client delay D. Figure 5.3 shows the weighted average server bandwidth usage of the six protocols with the default parameter settings except with D values of 1.1, 1.5, 10 and 100, in which case the batching delay parameter is 0.1, 0.5, 9 and 99 respectively. As the data request rate increases within the defined range, the bandwidth curve of every protocols has the same tendency of going upwards and eventually levelling out after the data request rate is high enough. From Figure 5.1 and Figure 5.3, as the maximum client delay increases, the highest server bandwidth usage for each of the protocols is reduced, the data request rate at which the server bandwidth usage flattens out becomes lower, and the crossover point at which the server bandwidth usage with SFN broadcast surpasses that with single-cell broadcast transmission shifts to a lower data request rate and lower server bandwidth. For the protocols using batching/cbd scheduling, the weighted average server bandwidth usage with the hybrid broadcast transmission scheme always closely matches the minimum of that with sing-cell broadcast or SFN broadcast. For the protocols using cyclic/cd,fft scheduling, only when the batching delay is equal to the data transmission time, as in Figure 5.1, does the weighted average server bandwidth usage with hybrid transmission always closely match the minimum of that with single-cell broadcasts or SFN broadcasts. When the batching delay is shorter than the data transmission time, for protocols using cyclic/cd,fft scheduling, the weighted average server bandwidth usage with hybrid broadcast slightly deviates from that with SFN broadcast over a range of data request rates immediately after the crossover point. When the batching delay is longer than the data transmission time, for protocols using cyclic/cd,fft scheduling, the weighted average server bandwidth usage with hybrid broadcast slightly deviates from that with single-cell broadcast transmission over a range of data request rates right before the crossover point. If the batching delay is much greater than the data transmission time, the performance difference in terms of the weighted average server

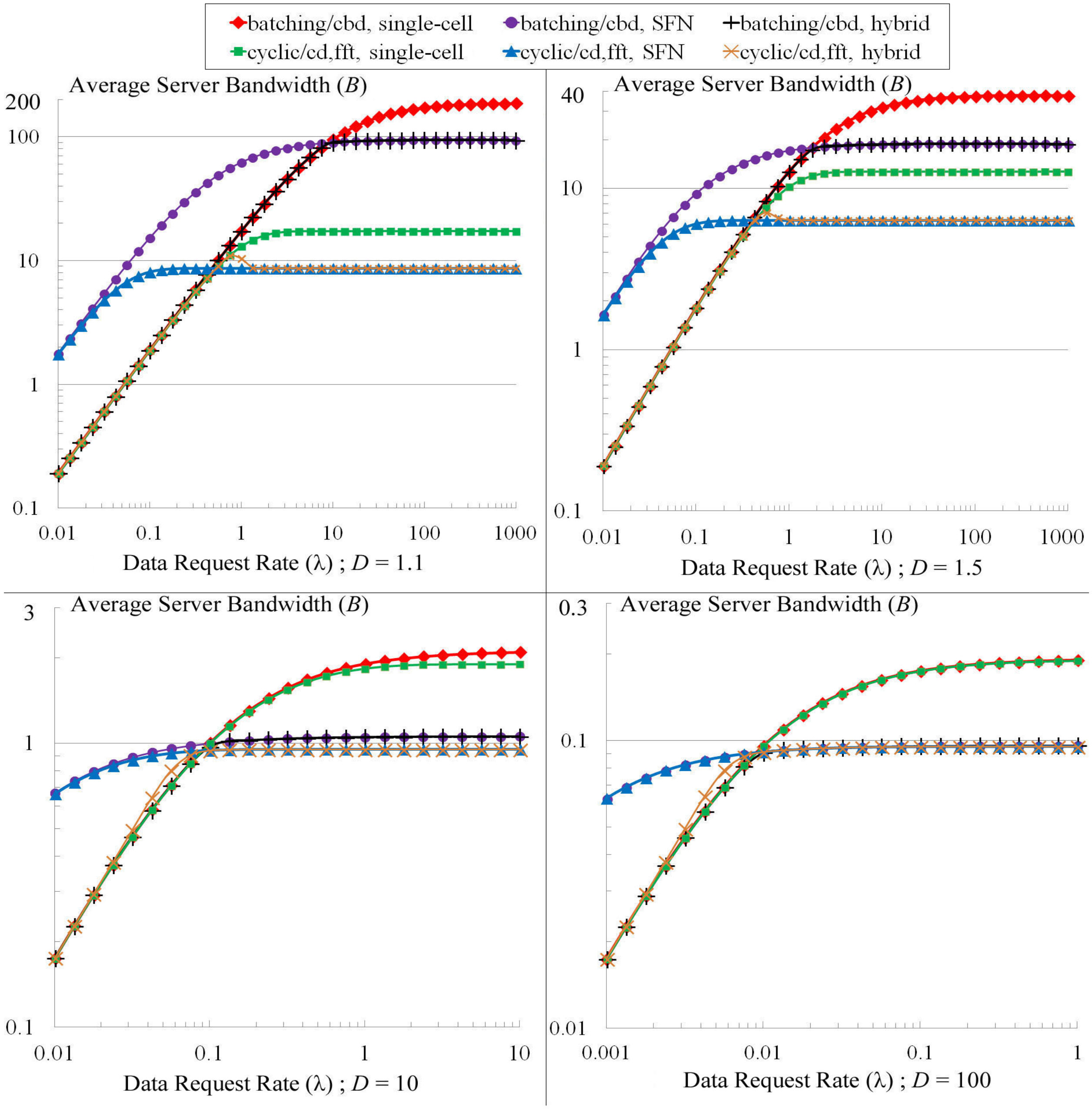


Figure 5.3: The weighted average server bandwidth usage of the six protocols under default parameter settings with $D = 1.1$, 1.5, 10 and 100.

bandwidth usage is negligible among the protocols using the same broadcast transmission scheme.

Figure 5.4 shows the average client delay of the six protocols with the default parameter setting except with D values of 1.1, 1.5, 10 and 100, in which case the batching delay parameter is 0.1, 0.5, 9 and 99 respectively. As the data request rate increases within the defined range, the average client delay of all protocols declines from the same value, and eventually stabilizes. The average client delay curve of the protocols with the same broadcast scheduling protocol converges at the same value. From Figure 5.2 and Figure 5.4, when

D increases, the lowest achievable average delay for each protocol is attained at lower data request rates and with a higher delay time. The lowest achievable average delay for the protocols using batching/cbd is lower than that for the protocols using cyclic/cd,fft, except when the batching delay is far longer than the data transmission time. In the case when D is 100, the lowest achievable average delay at high data request rates

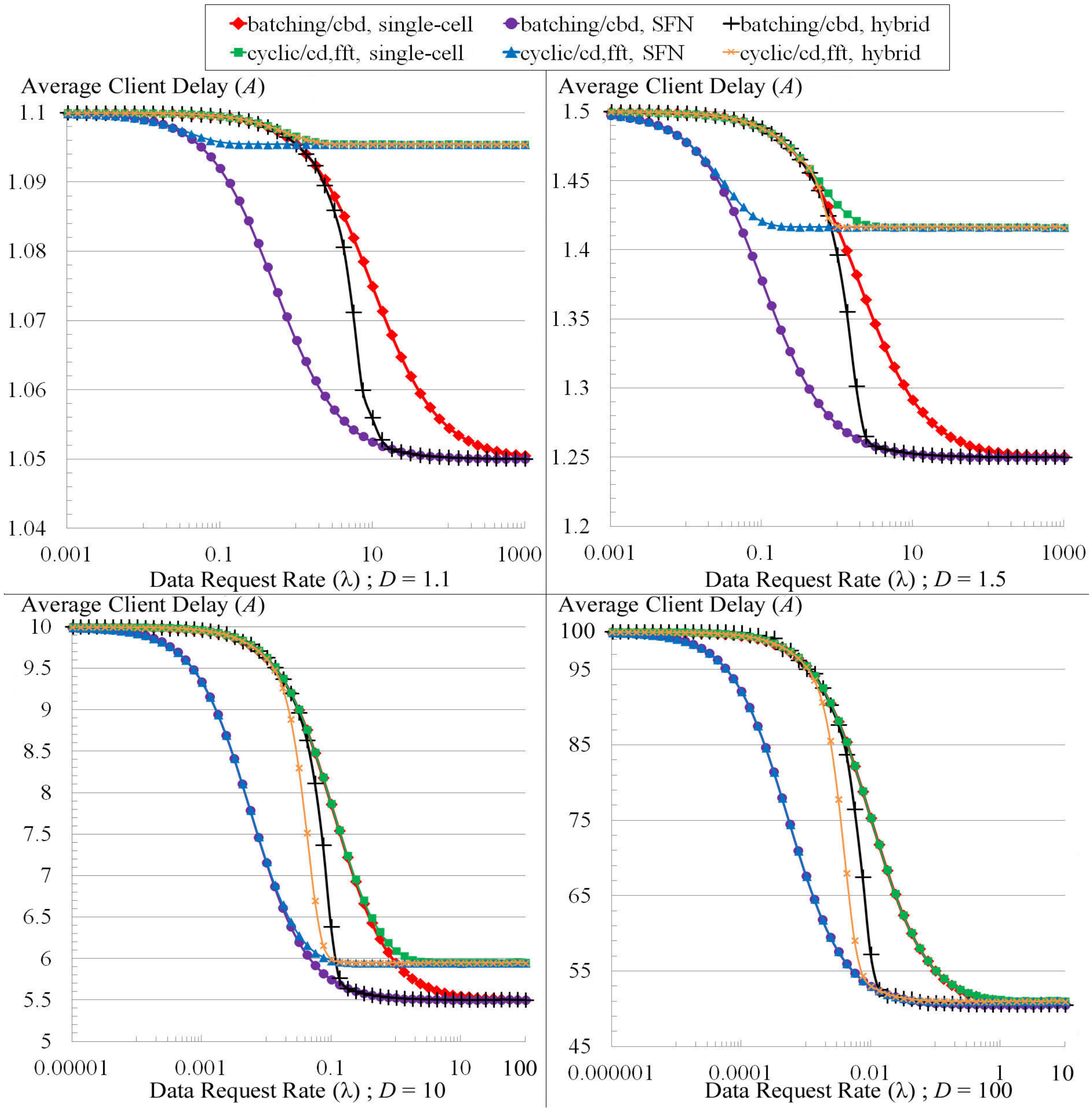


Figure 5.4: The average client delay of the six protocols under default parameter settings with D = 1.1, 1.5, 10 and 100

is close to the same for all of the protocols. Regardless of the value of D, the protocols using the same broadcast scheduling protocol, SFN broadcast generally yields lower average delay than single-cell broadcast, while hybrid broadcast yields intermediate average delay performance.

From the performance results with different D's, as the maximum client delay increases, the same broadcast protocol would have a lower weighted average server bandwidth usage, but a higher average delay, for the same request rate. When the batching delay is significantly shorter than or longer than the data transmission time, the protocol using cyclic/cd,fft scheduling with hybrid broadcast transmission does not always closely match the minimum weighted average server bandwidth among all the protocols. When the batching delay is much greater than the data transmission time, then a switch of the broadcast scheduling protocol would have little impact on the server bandwidth performance.

5.3.1 Accuracy of the Approximate Continuous-time State Transition Models for Hybrid Broadcast

In the previous chapter, two approximate continuous-time state transition models are proposed for each of the protocols using hybrid broadcast transmission. The difference between the two state transition models is that in the first state transition model there are multiple states in which SFN broadcasts occur and every SFN broadcast state has a different transition rate to state 0 in which there are no waiting requests, whereas in the second state transition model these SFN broadcast states are combined into one state. The two approximate continuous-time state transition models for the same broadcast scheduling protocol provide different estimations of the weighted average server bandwidth usage and the average client delay (in 4.20 & 4.21 and 4.22 & 4.23). To assess the accuracy of these models, experiment results from evaluating the models are compared to the simulation results. For each protocol, the analytic results from the models for the weighted average server bandwidth usage and the average client delay are compared to the corresponding simulation results. The approximate continuous-time state transition model whose results are closest to the simulation results is deemed to be more suitable for modelling hybrid broadcast transmission.

Figure 5.5 presents the percentage difference between the simulation results and the weighted average server bandwidth usage derived from each of the two continuous-time state transition models. Figure 5.6 presents the percentage difference between the simulation results and the average client delay derived from each of the two continuous-time state transition models. The experiments are carried out with D = 1.1, 2, 10 and 100. The comparison of the results shows that the analytic results from the continuous-time state transition models closely match the simulation results at fairly low data request rates. Once the data request rate has increased beyond a certain point, the differences between the analytic results and the simulation results increase until a peak point is reached. Then the analytic results slowly converge towards the simulation results. At fairly high data request rates, the analytic results remain close to the simulation results. The explanation for the close match between the analytic results and the simulation results at the boundary arrival rates is that either one of the two components of the continuous-time state transition models, which are the single-cell broadcast component and the SFN broadcast component, is accurate by itself when only one component plays the major part for broadcast. When both the single-cell broadcast component and the SFN broadcast component are used together for broadcast, then the continuous-time state transition models deviate from the true behaviour of the system, and this accounts for the poor accuracy for continuous-time state transition models at intermediate arrival rates. As D increases from 1.1 to 100, because the batching

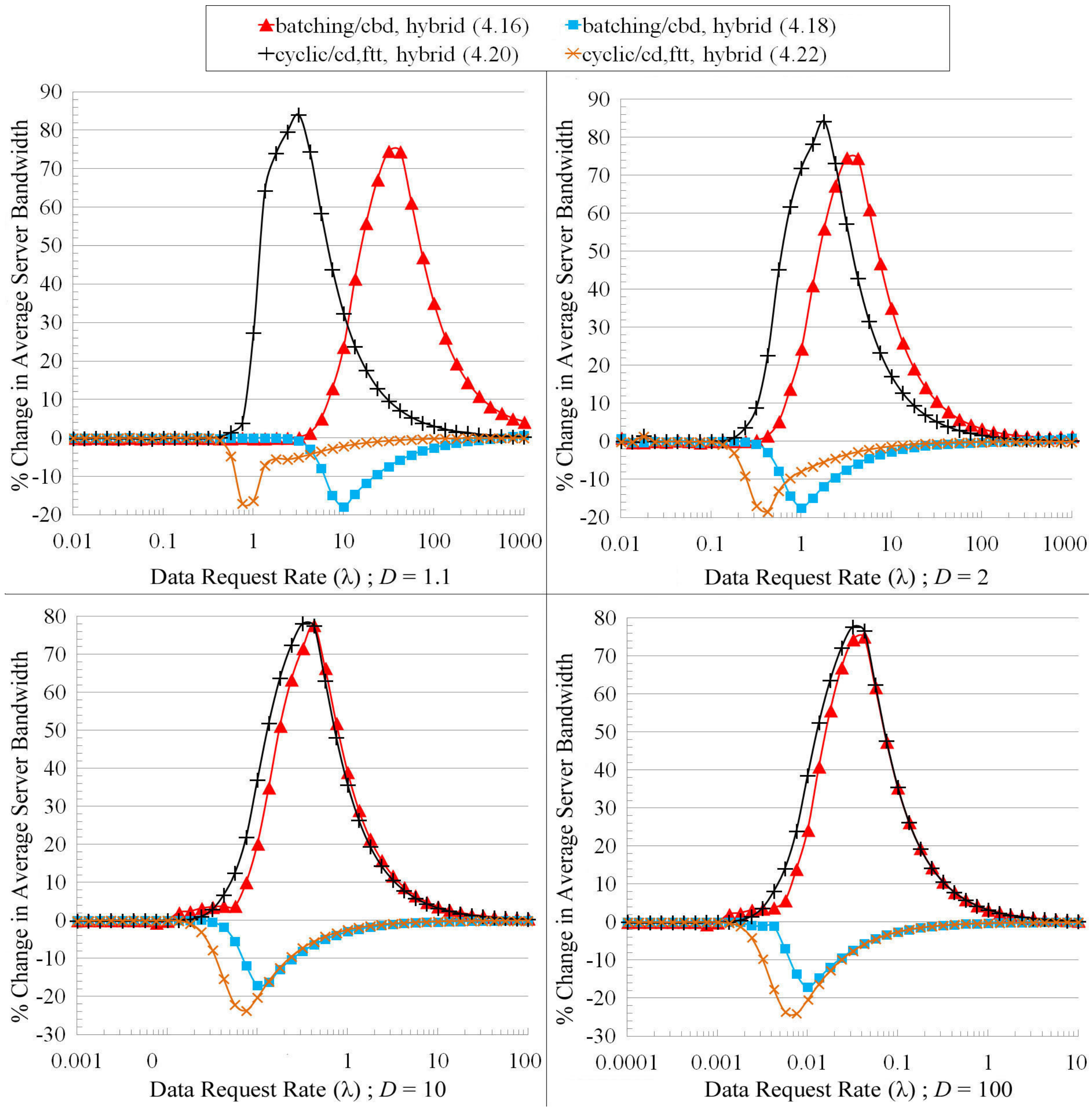


Figure 5.5: The percentage difference between the weighted average server bandwidth usage computed from each of the continuous-time state transition models relative to the simulation results with D = 1.1, 2, 10, 100.

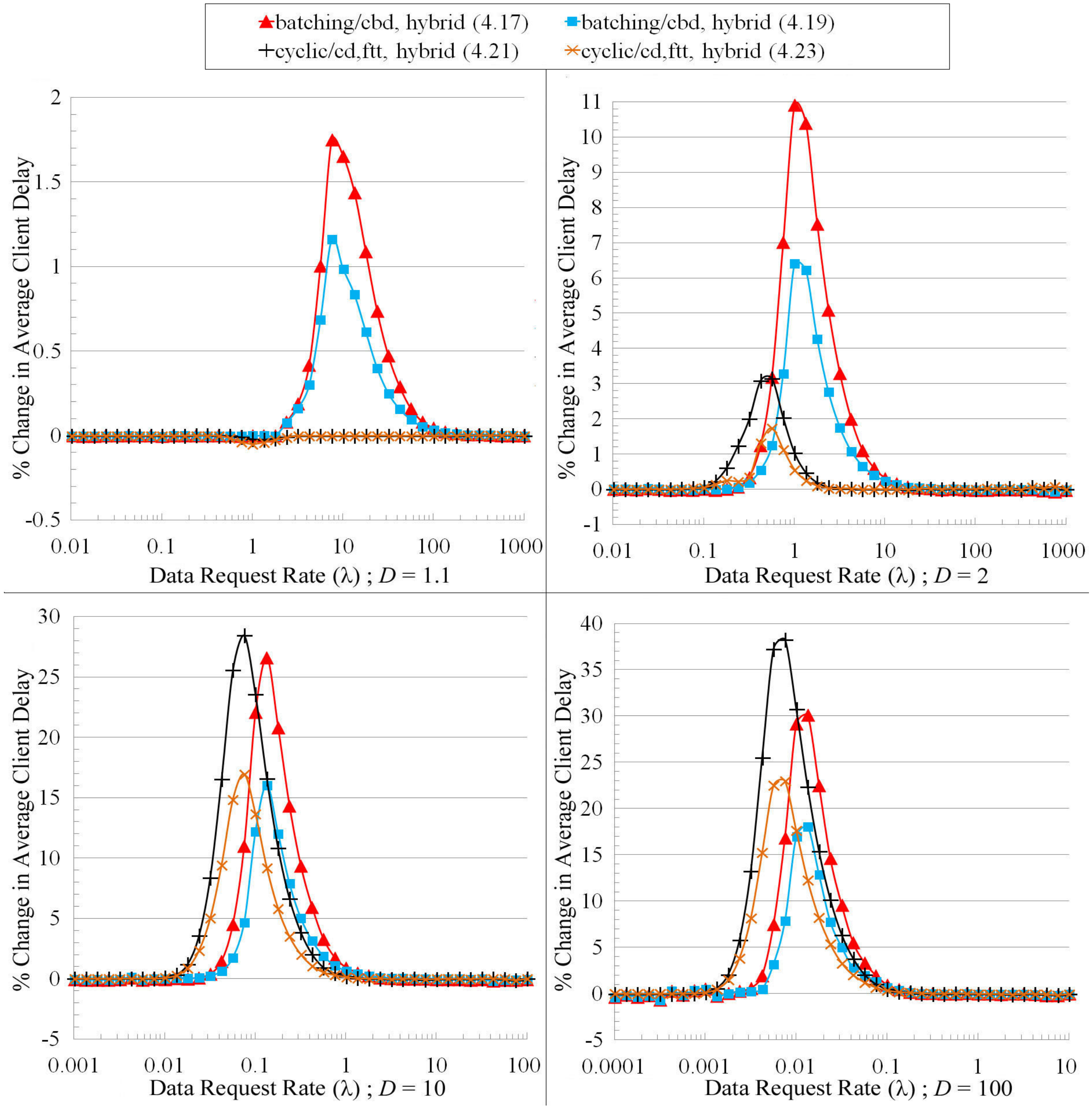


Figure 5.6: The percentage difference between the average client delay computed from each of the approximate continuous-time state transition models relative to the simulation results with $D = 1.1$, 2, 10, 100.

delay for accumulating enough number of data requests for the next SFN broadcast transmission becomes longer, the minimum data request for SFN broadcast transmissions to be continuous gets lower. That is why the observed range of data request rates gets narrower for larger D. For each of the broadcast scheduling protocols, the results from the second approximate continuous-time state transition model with only a single SFN broadcast state are closest to the simulation results than the results from the first model. Therefore, for performance analysis of the hybrid broadcast scheme, the model with only a single SFN broadcast state can be expected to be more accurate than the other proposed model that has multiple SFN broadcast states.

5.4 Results with D Chosen as Unit of Time (D=1)

In the previous experiments, the unit of data volume is picked to be the file size (L=1) and the unit of time is picked to be the time required for one complete transmission of the file (r=1). To provide a different perspective on performance, in this section the unit of time is defined to be the maximum client delay (D=1), and performance is assessed with different combinations of values for the size of the broadcast data file L and the data transmission rate r. Note that a change of units has no impact on actual performance, except for the units in which the performance metrics are measured. For example, given the input settings and output results from an experiment with variable D and constant L and r, the same input and output may be modified to provide the results for experiments with variable L and constant D and r, or the results for experiments with variable r and constant D and L. This conversion can be achieved by changing the time unit and/or the data volume unit, as summarized in Table 5.2. In Table 5.2, the input and output of case 1 corresponds to the input settings and results obtained from the initial set of simulation experiments. To obtain the corresponding inputs and results for cases 2, the value of D changes from val_D to the constant unit value, which can be interpreted as the related unit of time being increased by a factor of val_D from that in case 1. This change of time unit is applied to all the time-related factors in case 1 and results in the input and output of case 2. From case 2 to case 3, the value of r changes from val_D to the constant unit value, which can be interpreted as the unit of data volume being increased by a factor of val_D. This change of data volume unit is applied to all the data volume-related factors in case 2 and results in the input and output values in case 3.

From Table 5.2, the output of case 2 and case 3 may be claimed using the results of Figures 5.2 and 5.3 in which case val_D is 2. Figure 5.7 plots the corresponding results. In Figure 5.7, r and L, in turn, are changed to be 2 and 0.5 respectively, which corresponds to the case 2 and the case 3 in Table 5.2.

	Input						Output	
	L	r	D	λ	N	g	B	A
Case 1 with fixed values for L & r	1	1	val_D	val_λ	val_N	val_g	val_B	val_A
Case 2 with fixed values for L & D	1	val_D	1	$val_\lambda * val_D$	val_N	val_g	$val_B * val_D$	val_A/val_D
Case 3 with fixed values for r & D	$1/val_D$	1	1	$val_\lambda * val_D$	val_N	val_g	val_B	val_A/val_D

Table 5.2: The three cases that have equivalence relations with one another

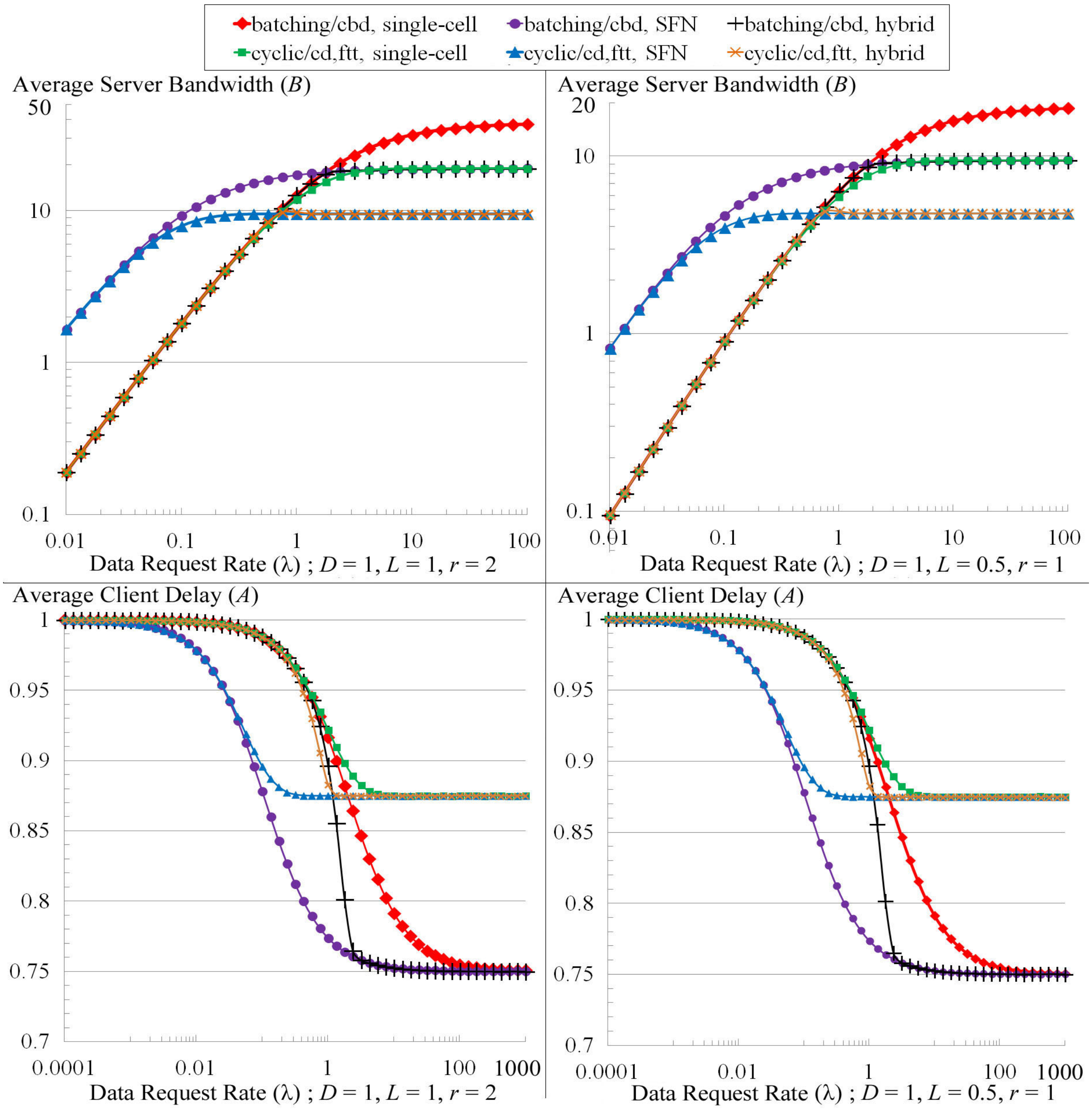


Figure 5.7: The weighted average server bandwidth usage and average client delay of the six protocols under default parameter settings with maximum client delay D=1

5.5 Results with Variable g

The quotient of the per-cell cost of an SFN broadcast divided by the cost of a single-cell broadcast, g, is a factor determined by the network design. For a network with 19 cells, the value for g should be in the range between 1/19 and 1. When g decreases within the defined value range, SFN broadcast becomes more attractive relative to single-cell broadcast transmission. With the hybrid broadcast transmission scheme, an increased likelihood of using SFN broadcast can be achieved by reducing the threshold value. Thus, the value of the hybrid broadcast threshold has to be adjusted when g changes. To demonstrate the impact of g on the hybrid broadcast transmission scheme, the protocols using hybrid broadcast are re-evaluated with various values for g, with the parameters D, L, r and N set at their default values.

Figure 5.8 presents the comparison of the weighted average server bandwidth usage with single-cell broadcast and hybrid broadcast with various g's. Two separate graphs are plotted from the collected bandwidth results for batching/cbd and cyclic/cd,fft scheduling. The input parameters D, L, r and N are assigned with the default values from Table 5.1, which the value for g is chosen as 0.1, 0.3, 0.5, 0.7 and 0.9. With any given g, T is calculated as the ceiling of g times N, which is a reasonable estimation for a good hybrid broadcast threshold. So the corresponding values for the threshold T are 2, 6, 10, 14 and 18. The data request rate ranges from 0.001 to 100. For both batching/cbd and cyclic/cd,fft scheduling, the single-cell broadcast and hybrid broadcast transmission schemes with variable g have the same weighted average server bandwidth usage at low data request rates, and the server bandwidth usage grows linearly with the increase in the data request rate. When the data request rate is high enough, the server bandwidth usage stabilize at different

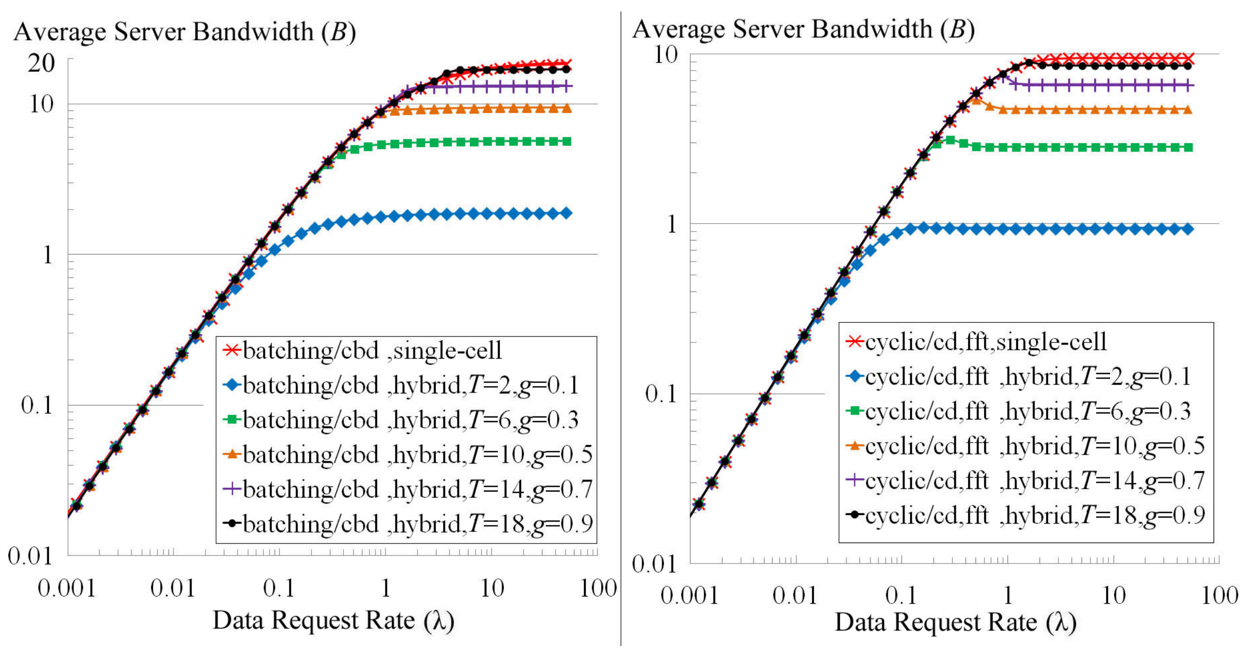


Figure 5.8: The weighted average server bandwidth usage using single-cell broadcasts and the hybrid broadcast transmission scheme under default parameter settings with g = 0.1, 0.3, 0.5, 0.7 and 0.9

levels. For the hybrid broadcast transmission schemes, as g changes from 0.1 to 0.9, the weighted average server bandwidth usage stabilizes at a higher value and at a higher data request rate. The highest weighted average server bandwidth usage with the hybrid broadcast transmission scheme with g=0.9 is slightly lower than that with the single-cell broadcasts. In conclusion, with frequent request arrivals, the server bandwidth performance of the hybrid broadcast transmission scheme improves as g decreases, and the weighted average server bandwidth usage with the hybrid broadcast transmission scheme is always less than that with single-cell broadcasts for any given g less than 1.

5.6 Results with Variable T

The value for the hybrid broadcast threshold is defined by the mobile broadcast system. Note that when the hybrid broadcast threshold T is 1, the hybrid broadcast scheme is the same as the SFN broadcast transmission scheme, and that the use of single-cell broadcast increases as T increases. Figure 5.9 presents the comparison of the weighted average server bandwidth usage with single-cell broadcasts, SFN broadcasts and the hybrid broadcast transmission scheme for various values of T. Two separate graphs are plotted from the results for the batching/cbd and cyclic/cd,fft scheduling protocols. The input parameters D, g, L, r and N are given the default values from Table 5.1, while the various values for T in the experiment are 2, 5, 10, 16 and 19. The per-cell data request rate ranges from 0.0005 to 50. With both batching/cbd and cyclic/cd

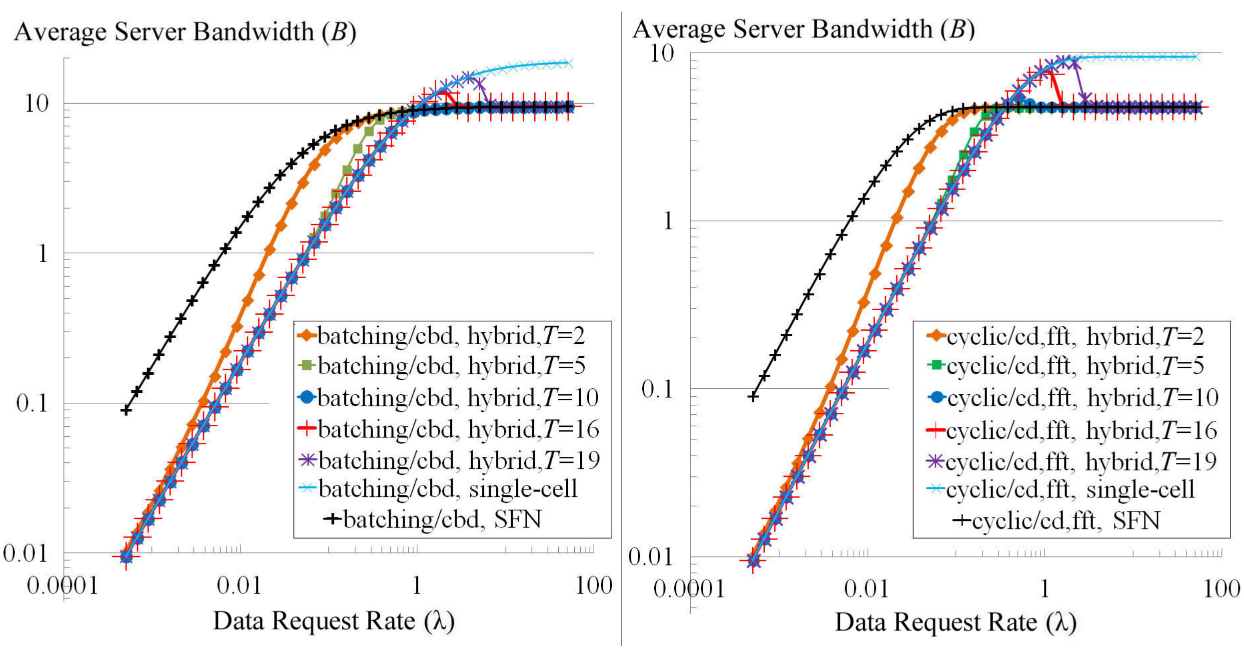


Figure 5.9: The weighted average server bandwidth usage using SFN broadcasts, single-cell broadcasts, and the hybrid broadcast transmission scheme under default parameter settings with different threshold parameter values

scheduling, as the data request rate increases from 0.0005 to 50, the weighted average server bandwidth usage increases, although not monotonically for the hybrid broadcast scheme, and eventually stabilizes. The highest weighted average server bandwidth usage is the same for all the cases except with the single-cell transmission scheme, in which the server bandwidth usage stabilizes at a higher value. Among all the values for the threshold T that are considered in this experiment, when T equals 10 the weighted average server bandwidth usage with the hybrid transmission scheme most closely matches the minimum of that with single-cell broadcast or SFN broadcast, for both batching/cbd and cyclic/cd,fft scheduling. When T is smaller than 10, the weighted average server bandwidth usage with hybrid broadcast significantly deviates from the minimum achieved with single-cell broadcast, just before the crossover point of the SFN and single-cell broadcast bandwidth curves. When T is greater than 10, the weighted average server bandwidth usage with the hybrid broadcast significantly deviates from the minimum, achieved with SFN broadcast, just after the crossover point of the SFN and single-cell broadcast bandwidth curves.

To compare the weighted average server bandwidth usage for hybrid broadcast with different T's, the root-mean-square deviation (RMSD) of the hybrid broadcast weighted average server bandwidth usage compared to the minimum of that with single-cell or SFN broadcast can be calculated and used as the benchmark. Table 5.3 presents the RMSD values for both batching/cbd and cyclic/cd,fft scheduling. These results use the default values for D, g, L, r and N from Table 5.1 and the values for the hybrid broadcast threshold T include all possible integer values between 2 and 19. Table 5.3 shows that the hybrid broadcast transmission scheme under the default parameter settings has the minimized weighted average server bandwidth usage when T=11 for batching/cbd scheduling, and when T=8 for cyclic/cd,fft scheduling. The optimal threshold value intuitively should be around g times N, since this is the number of cells at which an SFN broadcast is the same cost as that of multiple single-cell broadcasts. For the default parameter, $g \times N$=9.5. As T changes away from the optimal threshold value in either direction, the weighted average server bandwidth usage would increase for hybrid broadcast. When T reaches the boundary value 2 or 19, the weighted average server

Hybrid broadcast threshold T	2	3	4	5	6	7	8	9	10
RMSD with batching/cbd scheduling	1.613	1.262	0.976	0.733	0.528	0.354	0.210	0.102	0.042
RMSD with cyclic/cd,fft scheduling	0.865	0.581	0.382	0.239	0.138	0.070	0.040	0.059	0.108

Hybrid broadcast threshold T	11	12	13	14	15	16	17	18	19
RMSD with batching/cbd scheduling	0.040	0.053	0.109	0.220	0.369	0.564	0.787	1.066	1.485
RMSD with cyclic/cd,fft scheduling	0.173	0.260	0.354	0.473	0.584	0.731	0.854	1.029	1.234

Table 5.3: The root-mean-square deviation (RMSD) of the weighted average server bandwidth usage when using the hybrid broadcast transmission scheme for all possible values for T from the minimum of the server bandwidth usages when using SFN or single-cell broadcast transmission schemes under default parameter settings

bandwidth usage of hybrid broadcast has the most deviation from the minimum weighted server bandwidth usage.

The Figure 5.10 presents a comparison of the average client delay with single-cell broadcast, SFN broadcast, and hybrid broadcast, for various values of T. Two separate graphs are plotted from the results for batching/cbd and cyclic/cd,fft scheduling. The values for the hybrid broadcast threshold include 2, 10, 16 and 19. In the experiment, as the per-cell data request rate changes from 0.0001 to 1000, the average delay curves descend from the same starting point (equal to the maximum delay D) until they level out at same value (equal to $\Delta/2 + L/r$ for batching/cbd, and $[\Delta/(\Delta + L/r)] \times \Delta/2 + [(L/r)/(\Delta + L/r)] \times \Delta$ plus the file transmission time L/r for cyclic/cd,fft). For both batching/cbd and cyclic/cd,fft, the single-cell broadcast transmission scheme always gives the highest average delay and SFN broadcast the lowest average delay. As T decreases the average delay performance of the hybrid broadcast transmission scheme improves and its average delay curve deviates more from that of single-cell broadcast and becomes closer to that of SFN broadcast. When T equals 2, the average delay of the hybrid broadcast transmission scheme with batching/cbd scheduling becomes nearly the same as the average delay of SFN broadcast with batching/cbd scheduling but this is not the case when using cyclic/cd,fft scheduling. In conclusion, the hybrid broadcast transmission

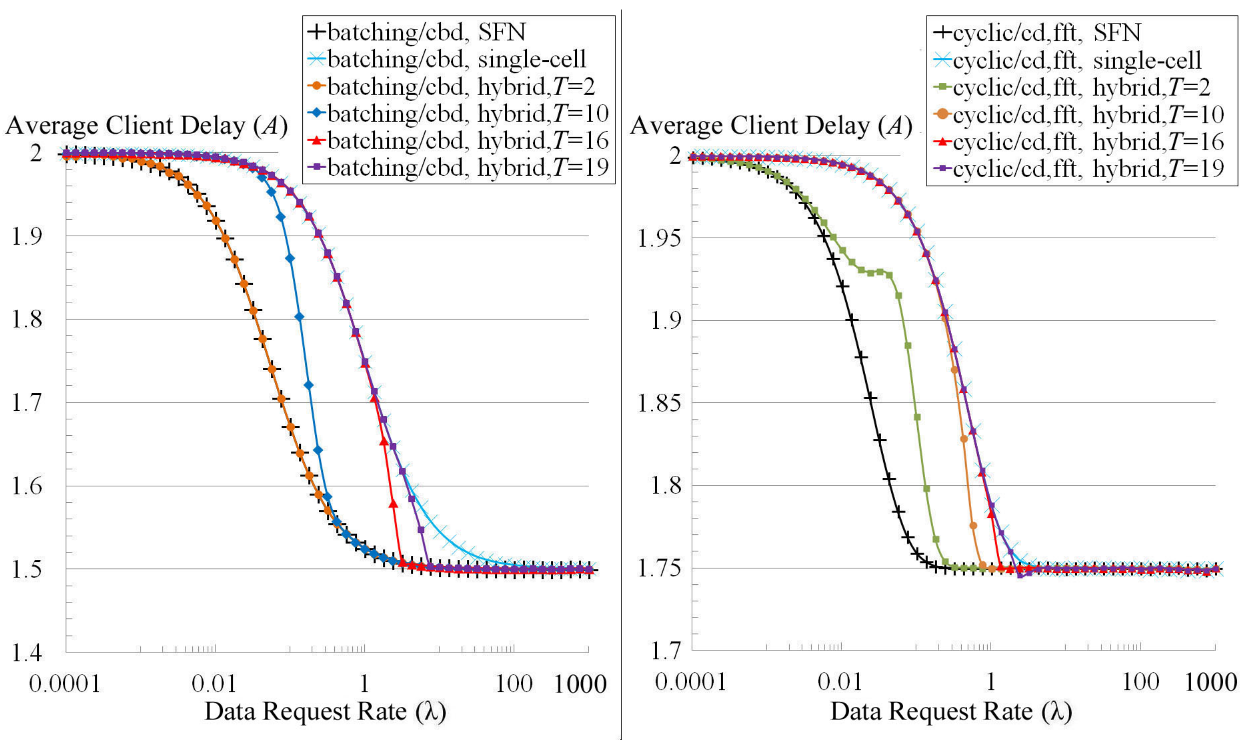


Figure 5.10: The average client delay using SFN broadcasts, single-cell broadcasts, and the hybrid broadcast transmission scheme under default parameter settings with different threshold parameter values

scheme is expected to have optimal weighted average server bandwidth usage when the hybrid broadcast threshold is set to be around the value of g time N. The change of threshold value in the hybrid broadcast scheme also has an impact on its average delay performance.

5.7 Results with Variable N

Figure 5.11 presents the comparison of the weighted average server bandwidth usage of the protocols using hybrid broadcast for various values of the number of cells N. Figure 5.12 presents the comparison of average client delay using hybrid broadcast for various values of N. In the experiment, the parameters D, g, L and r are given the default values from Table 5.1 and the different values for N are chosen as 11, 16, 19, 24, 29 and 34. The weighted average server bandwidth usage results for batching/cbd and cyclic/cd,fft scheduling are presented in two separate graphs in Figure 5.11. The average delay results for batching/cbd and cyclic/cd,fft scheduling are presented in two separate graphs in Figure 5.12. With any given N, the hybrid broadcast threshold T is chosen as the ceiling of g times N. For the hybrid broadcast transmission scheme with batching/cbd scheduling, as the data request rate increases from 0.001 to 100, the weighted average server bandwidth usage at first grows linearly until it reaches a peak value. It then stabilizes at the value LgN/Δ for batching/cbd scheduling, and $LgN/(\Delta + L/r)$ for cyclic/cd scheduling. Note that with cyclic/cd scheduling, the weighted average server bandwidth usage experiences a slight slump just before it levels out. As the value for N increases, the weighted average server bandwidth usage increases at any data

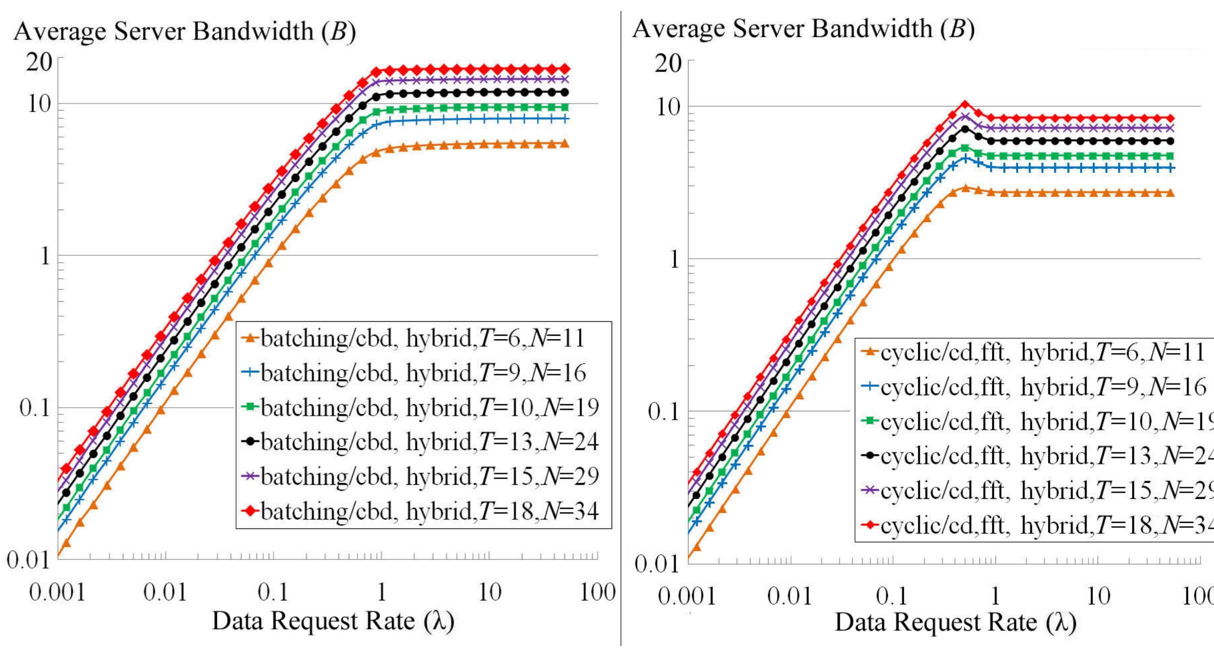


Figure 5.11: The weighted average server bandwidth using the hybrid broadcast transmission scheme under default parameter settings with N = 11, 16, 19 ,24, 29 and 34.

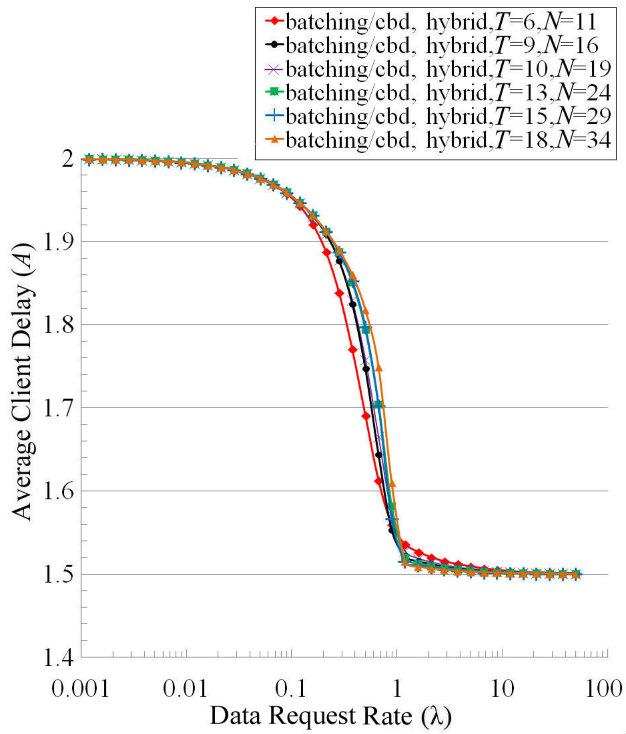


Figure 5.12: The average client delay using the hybrid broadcast transmission scheme under default parameter settings with $N = 11$, 16, 19 ,24, 29 and 34.

request rate. Within the defined value range of the data request rates, the average delay curves for either batching/cbd or cyclic/cd scheduling all begin descending from the same starting point and eventually level out at the same stabilized average delay at high data request rates (equal to $\Delta/2 + L/r$ for batching/cbd, and $[\Delta/(\Delta + L/r)] \times \Delta/2 + [L/r/(\Delta + L/r)] \times \Delta + L/r$ for cyclic/cd,fft). When the number of cells in the MBSFN area varies, the average delay for the hybrid broadcast transmission scheme with the same broadcast scheduling protocol remains largely the same and this results from the precondition that the maximum delay, and specifically Δ, L and r are defined to be fixed. In conclusion, with the hybrid broadcast transmission scheme the increases in the number of cells in the MBSFN area would directly raise the weighted average server bandwidth usage for data service, but only have minimal impact on the average client delay.

5.8 Summary

In this chapter, performance results are presented for six different mobile broadcast protocols. Simulation is used to assess the performance, with results that match those of the exact analytic models for the single-cell and SFN broadcast transmission schemes, and that are used to evaluate the accuracy of the approximate analytic models for the hybrid broadcast transmission scheme. In each experiment, the data request rate, which is chosen as the same value for every cell, varies over a wide range. The performance metrics are the weighted average server bandwidth usage and the average client delay. The protocols are first evaluated under default parameter settings. It is shown that the single-cell broadcast transmission scheme has a lower weighted average server bandwidth usage than the SFN broadcast transmission scheme when the data request rate is low relative to the data transmission rate. The SFN broadcast transmission scheme has a lower weighted average server bandwidth usage than the single-cell broadcast transmission scheme when the data request rate is high relative to the data transmission rate. The weighted average server bandwidth usage with the hybrid broadcast transmission scheme is close to the minimum of that with SFN or single-cell broadcast transmission. The SFN broadcast transmission scheme generally yield lower average delay than the single-cell broadcast transmission scheme or the hybrid broadcast transmission scheme, for the same batching delay parameter and maximum client delay.

The input parameters which would significantly affect the mobile broadcast performance are then varied one at a time in the experiments. These input parameters include the maximum client delay, the quotient of the per-cell cost of an SFN broadcast divided by the cost of a single-cell broadcast, the hybrid broadcast threshold and the number of cells in the MBSFN area. The file size is chosen as the unit of data volume (L=1) and the time required for one complete transmission of the file is chosen as the unit of time (r=1). Simulation experiments are also carried out with different data transmission rate or size of the broadcast data file, in which cases the maximum client delay is chosen as the unit of time.

Results for the weighted average server bandwidth usage and the average delay, with the default parameter settings except for various alternative values for D, are used as the benchmark for assessing the accuracy of the approximate continuous-time state transition models for the hybrid broadcast transmission scheme. From the experiment results, the models with a single SFN broadcast state are shown to be relatively more accurate.

Chapter 6

Summary and Conclusion

High data rate and and efficient use of radio spectrum are advanced features of the 4G mobile technology. The performance boost and new capabilities in LTE 4G presents an opportunity for use of mobile broadcast, which may provide performance benefit in certain use case scenarios. In light of the ever-growing data traffic burden in the mobile network, mobile broadcast may serve as an alternative approach to the point-to-point unicast transmission for large-scale data dissemination. Besides the single-cell broadcast transmission, LTE 4G also enables SFN broadcast transmission, in which data broadcasting is synchronized in a number of cells referred to as the MBSFN area. By introducing an efficient mechanism for switching between the two broadcast transmission schemes, single-cell broadcast transmission and SFN broadcast transmission may be used interchangeably for data service in the same mobile network.

6.1 Thesis Summary

One current challenge that comes along with mobile broadcast is the need for a suitable multi-cell on-demand broadcast protocol. There are on-demand broadcast scheduling protocols from previous work that are designed for data service in the single-cell network. The batching/cbd protocol, which is from the previous work, can be adapted for the multi-cell data service. Based on the previously proposed combined batching and cyclic broadcast protocol, cyclic/cd,bot, the cyclic/cd,fft protocol is newly proposed as the other candidate scheduling protocol for the multi-cell data service. The three possible broadcast transmission schemes enabled in LTE 4G include the single-cell broadcast transmission, the SFN broadcast transmission, and the hybrid broadcast transmission which combines the use of single-cell and SFN broadcast transmission in the same mobile broadcast network. A mobile broadcast protocol is made up of a multi-cell broadcast scheduling protocol and a broadcast transmission scheme. From the two multi-cell broadcast scheduling protocols and the three broadcast transmission schemes, six candidate mobile broadcast protocols are proposed in this research for mobile data service.

The mobile broadcast protocols are evaluated through development of analytic models and through simulation experiments. The performance metrics are the weighted average server bandwidth usage, the average client delay, and the maximum client delay. The maximum client delay for all of the protocols is $\Delta + L/r$, where Δ is the batching delay protocol parameter, L is the file size, and r is the transmission rate. Other important parameters include the quotient of the per-cell cost of an SFN broadcast divided by the cost a single-cell broadcast (g), the data request rate in each cell i (λ_i), the hybrid broadcast threshold (T) and the number of cells in the mobile broadcast network (N).

Given the assumptions, exact analytic models can be derived for single-cell and SFN broadcast transmission, for both batching/cbd and cyclic/cd,fft scheduling, yielding equations for the weighted average server bandwidth usage and the average client delay. For the hybrid broadcast transmission scheme, only approximate continuous-time state transition models could be derived. Two different approximate continuous-time state transition models were derived for each of batching/cbd and cyclic/cd,fft scheduling.

Simulation programs were developed for the six mobile broadcast protocols. In the simulation experiments, the performance of the protocols as measured by the weighted average server bandwidth usage and the average client delay was assessed as the data request rate per cell, varies over a wide range. First performance is assessed under default parameter settings. The single-cell broadcast transmission scheme has a lower weighted average server bandwidth usage than the SFN broadcast transmission scheme when the data request rate is low relative to the data transmission rate. The SFN broadcast transmission scheme has a lower weighted average server bandwidth usage than the single-cell broadcast transmission scheme when the data request rate is high relative to the data transmission rate. The weighted average server bandwidth usage with the hybrid broadcast transmission scheme is close to the minimum of that with SFN or single-cell broadcast transmission. The SFN broadcast transmission scheme generally yields lower average delay than the single-cell broadcast transmission scheme or the hybrid broadcast transmission scheme.

In subsequent experiments, the various input parameters are varied one at a time, so that their impact on data transmission performance can be assessed. These input parameters are the maximum client delay, the quotient of the per-cell cost of an SFN broadcast divided by the cost of a single-cell broadcast, the hybrid broadcast threshold and the number of cells in the MBSFN area. The choice of data transmission rate and the size of the broadcast data file server only to fix the units of time and data volume. Additional experiments are further carried out with the unit of time chosen as the maximum client delay, and with various values of the data transmission rate and the size of the broadcast data file.

Simulation experiments are also used to assess the accuracy of the approximate continuous-time state transition models for the hybrid broadcast transmission scheme. The simulation results for the weighted average server bandwidth usage and the average client delay are regarded as the benchmark. From the comparison of the results, the approximate continuous-time state transition model with only a single SFN broadcast state is found to be more accurate for representing hybrid broadcast transmission performance because its estimated results in both weighted average server bandwidth usage and average client delay are closer to the results from the simulation experiments.

6.2 Thesis Contributions

The main contributions of this thesis are as follows:

- A hybrid broadcast transmission scheme which combines the use of single-cell broadcast transmission and SFN broadcast transmission is designed for mobile broadcast.

- A multi-cell broadcast scheduling protocol, the cyclic/cd,fft protocol, is designed for mobile data service.

- Six mobile broadcast protocols are proposed based on the three broadcast transmission schemes enabled by LTE 4G and two multi-cell broadcast scheduling protocols.

- Analytic performance models were developed for each of these six protocols, yielding equations for calculating the weighted average server bandwidth usage, the average client delay and the maximum client delay. The analytic models with the single-cell broadcast transmission or SFN broadcast transmission are exact, given the assumptions. For the hybrid broadcast transmission scheme, two approximate continuous-time state transition models are devised for each of batching/cbd and cyclic/cd,fft. Later by comparing the results from different estimations and the corresponding simulation results, the approximate continuous-time state transition model for each scheduling protocol with the most accurate estimations of the average server bandwidth and the average client delay is identified.

- Simulation protocols are developed for the six mobile protocols. The simulation experiments assess performance as measured by the weighted average server bandwidth usage and the average client delay under default parameter settings, as well as when the input parameters are varied one at a time. These input parameters include the maximum client delay, the quotient of the per-cell cost of an SFN broadcast divided by the cost of a single-cell broadcast, the hybrid broadcast threshold and the number of cells in the MBSFN area. Experiments are also carried out with various values of the data transmission rate and the size of the broadcast data file. The simulation results show that the hybrid broadcast transmission scheme together with the cyclic/cd,fft protocol provides the best weighted average server bandwidth usage and the SFN broadcast transmission scheme together with the batching/cbd protocol provides the best average delay performance for a given batching delay parameter and maximum client delay.

6.3 Future Work

In this thesis, simulation experiments are carried out with the same data request rate in every cell. In the future, the impact of heterogeneity of data request rates in different cells should be evaluated. In the real-world mobile network, it is common-place for the user devices to move between different cells even during a broadcast transmission. As a result, the impact of the variability of the request rate in each cell could be evaluated, and the constantly changing data request rate in every cell may be considered.

For both the analysis and the simulation experiment in this research, the MBSFN area is the entire mobile broadcast network. In mobile broadcast protocols using SFN or the hybrid broadcast transmission scheme, the size of the MBSFN area could be adjustable and the MBSFN area could consist of a subset of cells in the mobile broadcast network. Also the on-going development of LTE technology, some advanced features of the newest LTE release may be specifically adapted for mobile broadcast applications in the future studies.

The power consumption of components of the mobile broadcast system may be further considered for performance evaluation besides the weighted average server bandwidth usage and the average client delay. As the area of the mobile broadcast network expands, the required power consumption would rise accordingly. In the mobile broadcast network whose size is unchanged, the relation between the required power consumption and the data request rate could be evaluated. The mobile protocol with the minimized power consumption may also be identified.

Bibliography

[1] M. F. L. Abdullah and A. Z. Yonis. Performance of LTE Release 8 and Release 10 in Wireless Communications. In *Proc. 2012 International Conference on Cyber Security, Cyber Warfare and Digital Forensic (CyberSec)*, pages 236–241, Kuala Lumpur, Malaysia, Jun. 2012.

[2] S. Acharya and S. Muthukrishnan. Scheduling On-demand Broadcasts: New Metrics and Algorithms. In *Proc. 4th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom '98)*, pages 43–54, Dallas, TX, Oct. 1998.

[3] D. Aksoy and M. Franklin. RxW: A Scheduling Approach for Large-Scale On-Demand Data Broadcast. *IEEE/ACM Transactions on Networking (ToN)*, 7(6):846–860, Dec. 1999.

[4] I. F. Akyildiz, D. M. Gutierrez-Estevez, and E. C. Reyes. The Evolution to 4G Cellular Systems LTE-Advanced. *Physical Communication*, 3(4):217–244, Dec. 2010.

[5] A. Alexiou, C. Bouras, V. Kokkinos, and G. Tsichritzis. Communication Cost Analysis of MBSFN in LTE. In *Proc. IEEE 21st International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 1366–1371, Instanbul, Turkey, Sept. 2010.

[6] A. Alexiou, C. Bouras, V. Kokkinos, and G. Tsichritzis. Performance Evaluation of LTE for MBSFN Transmissions. *Wireless Networks*, 18(3):227–240, Apr. 2012.

[7] K. C. Almeroth, M. H. Ammar, and F. Zongming. Scalable Delivery of Web Pages Using Cyclic Best-effort Multicast. In *Proc. 7th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '98)*, pages 1214–1221, San Francisco, CA, Mar./Apr. 1998.

[8] D. Astély, E. Dahlman, G. Fodor, S. Parkvall, and J. Sachs. LTE Release 12 and Beyond. *IEEE Communication Magazine*, 51(7):154–160, Jul. 2013.

[9] D. Astély, E. Dahlman, A. Furuskär, Y. Jading, M. Lindstrőm, and S. Parkvall. LTE: the Evolution of Mobile Broadband. *IEEE Communications Magazine*, 47(4):44–51, Apr. 2009.

[10] M. Bakhuizen and U. Horn. Mobile Broadcast/Multicast in Mobile Networks. *Ericsson Review*, 1, 2005.

[11] S. Bhattacharyya, J. F. Kurose, D. Towlsey, and R. Nagarajan. Efficient Rate-controlled Bulk Data Transfer Using Multiple Multicast Groups. In *Proc. 7th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '98)*, pages 1172–1179, San Francisco, CA, Mar./Apr. 1998.

[12] Y. Birk and D. Crupnicoff. A Multicast Transmission Schedule for Scalable Multirate Distribution of Bulk Data Using Nonscalable Erasure-correcting Codes. In *Proc. 22nd Annual Joint Conference of the IEEE Computer and Communications (INFOCOM 2003)*, pages 1033–1043, San Francisco, CA, Mar./Apr. 2003.

[13] J. W. Byers, M. Luby, M. Mitzenmacher, and A. Rege. A Digital Fountain Approach to Reliable Distribution of Bulk Data. In *Proc. ACM 1998 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication (SIGCOMM '98)*, pages 56–67, Vancouver, Canada, Sept. 1998.

[14] J. Calabuig, J. F. Monserrat, D. Martn-Sacristn, and J. Olmos. Comparison of Multicast/Broadcast Services in Long Term Evolution Advanced and IEEE 802.16m Networks. *Wireless Communications and Mobile Computing*, 14(7):717–728, May 2014.

[15] N. Carlsson, D. L. Eager, and M. K. Vernon. Multicast Protocols for Scalable On-Demand Download. *Performance Evaluation*, 63(8/9):864–891, Oct. 2006.

[16] N. Cornillet, M. Crussiere, and J. Helard. On the Hybrid Use of Unicast/Broadcast Networks under Energy Criterion. In *Proc. IEEE 23rd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 1256–1261, Sydney, Australia, Sept. 2012.

[17] N. Cornillet, M. Crussiere, and J. Helard. Optimization of the Energy Efficiency of a Hybrid Broadcast/Unicast Network. In *Proc. 2013 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, pages 39–44, Shanghai, China, Apr. 2013.

[18] M. Dehghani and K. Arshad. LTE-advanced, and the Way Forward. In *Proc. 14th International Symposium on Communications and Information Technologies (ISCIT)*, pages 466–469, Incheon, South Korea, Sept. 2014.

[19] J. Erman and K. K. Ramakrishnan. Understanding the Super-sized Traffic of the Super Bowl. In *Proc. 2013 Conference on Internet Measurement Conference (IMC '13)*, pages 353–360, Barcelona, Spain, Oct. 2013.

[20] A. Finamore, M. Mellia, Z. Gilani, K. Papagiannaki, V. Erramilli, and Y. Grunenberger. Is There a Case for Mobile Phone Content Pre-staging? In *Proc. 9th ACM Conference on Emerging Networking Experiments and Technologies (CoNEXT '13)*, pages 321–326, Santa Barbara, CA, Dec. 2013.

[21] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas. LTE-advanced Next-generation Wireless Broadband Technology. *IEEE Wireless Communications*, 17(3):10–22, Jun. 2010.

[22] M. Gruber and D. Zeller. Multimedia Broadcast Multicast Service: New Transmission Schemes and Related Challenges. *IEEE Communication Magazine*, 49(12):176–181, Dec. 2011.

[23] F. Hartung, U. Horn, J. Huschke, M. Kampmann, T. Lohmar, and M. Lundevall. Delivery of Broadcast Services in 3G Networks. *IEEE Transactions on Broadcasting*, 53(1):188–199, Mar. 2007.

[24] C. Hoymann, D. Larsson, H. Koorapaty, and J. Cheng. A Lean Carrier for LTE. *IEEE Communications Magazine*, 51(2):74–80, Feb. 2013.

[25] R. Hwang, C. Huang, C. Lin, and C. Chung. Context-aware Multimedia Broadcast and Multicast Service Area Planning in 4G Networks. *Computer Communications*, Feb. 2015.

[26] A. Ibrahim and N. Seman. The Investigation of MBMS for Delivering Multimedia Contents in Single Frequency Network. In *Proc. IEEE Student Conference on Research and Development (SCOReD)*, pages 214–217, Pulau Pinang, Malaysia, Dec. 2012.

[27] A. Ibrahim, N. Seman, and C. Y. Leow. Single Frequency Network Evaluations For Multimedia Broadcast and Multicast Service (MBMS) Implementation In Outdoor and Indoor Area. *Jurnal Teknologi*, 63(3):6166, Apr. 2013.

[28] A. Ibrahim, N. Seman, and C. Y. (B.) Leow. The Investigation of MBMS over Single Frequency Network in Outdoor Urban Area. In *Proc. IEEE International RF and Microwave Conference (RFM)*, pages 125–128, Penang, Malaysia, Dec. 2013.

[29] D. Lecompte and F. Gabin. Evolved Multimedia Broadcast/Multicast Service(eMBMS) in LTE-Advanced: Overview and Rel-11 Enhancements. *IEEE Communication Magazine*, 50(11):68–74, Nov. 2012.

[30] G. Lee and S. Lo. Broadcast Data Allocation for Efficient Access of Multiple Data Items in Mobile Environments. *Mobile Networks and Applications*, 8(4):365–375, Aug. 2003.

[31] K. Liu and V. C. S. Lee. Simulation Studies on Scheduling Requests for Multiple Data Items in On-demand Broadcast Environments. *Performance Evaluation*, 66(7):368–379, Feb. 2009.

[32] K. Liu, V. C. S. Lee, and K. R. P. H. Leung. Data Scheduling for Multi-item Requests in Multi-channel On-demand Broadcast Environments. In *Proc. 7th ACM International Workshop on Data Engineering for Wireless and Mobile Access (MobiDE '08)*, pages 47–54, New York, NY, 2008.

[33] T. Lohmar, M. Slssingar, V. Kenehan, and S. Puustinen. Delivering Content with LTE Broadcast. *Ericsson Review*, 1(11):2–8, Feb. 2013.

[34] J. F. Monserrat, J. Calabuig, A. Fernndez-Aguilella, and D. Gmez-Barquero. Joint Delivery of Unicast and E-MBMS Services in LTE Networks. *IEEE Transactions on Broadcasting*, Apr. 2012.

[35] J. Na, D. Kim, J. Moon, S. Lee, and Y. Shin. A novel single-cell transmission scheme for multicast services in LTE-advanced. In *Proc. International Conference on Information Networking (ICOIN)*, pages 186–191, Phuket, Thailand, Feb. 2014.

[36] N. Nguyen and C. Bonnet. Service Continuity for eMBMS in LTE/LTE-Advanced Network: Standard Analysis and Supplement. In *Proc. IEEE 11th Consumer Communications and Networking Conference (CCNC)*, pages 219–224, Las Vegas, NV, Jan. 2014.

[37] S. Park and J. Kim. Effect of LTE Service Adoption on Mobile Data Traffic in Korea. In *Proc. 7th International Conference on Computing and Convergence Technology (ICCCT)*, pages 1197–1200, Seoul, South Korea, Dec. 2012.

[38] S. Parkvall, A. Furuskar, and E. Dahlman. Evolution of LTE toward IMT-advanced. *IEEE Communications Magazine*, 49(2):84–91, Feb. 2011.

[39] M. Phan and J. Huschke. Adaptive Point-to-multipoint Transmission for Multimedia Broadcast Multicast Services in LTE. *Computing Research Repository (CORR)*, Jul. 2009.

[40] L. Rong, O. B. Haddada, and S.-E Elayoubi. Analytical Analysis of the Coverage of a MBSFN OFDMA Network. In *Proc. IEEE 2008 Global Telecommunications Conference (GLOBECOM 2008)*, pages 1–5, New Orleans, LA, Nov./Dec. 2008.

[41] R. Sachan, N. Saxena, and A. Roy. An Efficient Hybrid Scheduling Scheme for Impatience User in eMBMS over LTE. In *Proc. 2013 International Conference on Computer Communication and Informatics (ICCCI)*, pages 1–5, Coimbatore, India, Jan. 2013.

[42] J. Santos, D. Gomes, S. Sargento, R. L. Aguiar, N. Baker, M. Zafar, and A. Ikram. Multicast/Broadcast Network Convergence in Next Generation Mobile Networks. *Computer Networks: The International Journal of Computer and Telecommunications Networking*, 52(1):228–247, Jan. 2008.

[43] Y. Sheng, X. Guo, M. Peng, and W. Wang. Analysis of Novel User Detection Scheme Based on Polling for E-MBMS Networks. In *Proc. IEEE 68th Vehicular Technology Conference (VTC)*, pages 1–5, Calgary, Canada, Sept. 2008.

[44] K. Stathatos, N. Roussopoulos, and J.S. Baras. Adaptive Data Broadcast in Hybrid Networks. In *Proc. 23rd International Conference on Very Large Data Bases (VLDB '97)*, pages 326–335, Athens, Greece, 1997.

[45] X. Sun, C. Wu, and Z. Zeng. Dynamic Power Saving Mechanism Joint Delivery of Unicast and E-MBMS Services in LTE MBSFN Networks. In *Proc. 16th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 189–194, The Hague, Netherlands, Oct. 2013.

[46] A. Urie, A. N. Rudrapatna, C. Raman, and J. Hanriot. Evolved Multimedia Broadcast Multicast Service in LTE: An Assessment of System Performance Under Realistic Radio Network Engineering Conditions. *Bell Labs Technical Journal*, 18(2):57–76, Sept. 2013.

[47] G. K. Walker, J. Wang, C. Lo, X. Zhang, and G. Bao. Relationship Between LTE Broadcast/eMBMS and Next Generation Broadcast Television. *IEEE Transactions on Broadcasting*, 60(2):185–192, May 2014.

[48] J. L. Wolf, M. S. Squillante, J. J. Turek, P. S. Yu, and J. Sethuraman. Scheduling Algorithms for the Broadcast Delivery of Digital Products. *IEEE Transactions on Knowledge and Data Engineering*, 13(5):721–741, Sept./Oct. 2001.

[49] J. W. Wong. Broadcast Delivery. In *Proc. IEEE,76(12)*, pages 1566–1577, Dec. 1988.

[50] J. W. Wong and M. H. Ammar. Analysis of Broadcast Delivery in a Videotex System. *IEEE Transactions on Computers*, C-34(9):863–866, Sept. 1985.

[51] C. Zhang, S. L. Ariyavisitaku, and M. Tao. LTE-Advanced and 4G Wireless Communications: Part 2. *IEEE Communication Magazine*, 50(6):26, Jun. 2012.